

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 9432372

Some consequences of semantic externalism

Preti, Consuelo, Ph.D.

City University of New York, 1994

Copyright ©1994 by Preti, Consuelo. All rights reserved.

U·M·I

**300 N. Zeeb Rd.
Ann Arbor, MI 48106**

11

SOME CONSEQUENCES OF SEMANTIC EXTERNALISM

by

CONSUELO PRETI

A dissertation submitted to the Graduate Faculty in
Philosophy in partial fulfillment of the requirements for
the degree of Doctor of Philosophy, The City University of
New York.

1994

copyright 1994

CONSUELO PRETI

All Rights reserved

This manuscript has been read and accepted for the Graduate Faculty in Philosophy in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

MARCH 16, 1994
Date

March 16, 1994
Date

William James Earle
Chair of Examining Committee

Richard Marshall
Executive Officer

Jerry Fodor

Stephen Schiffer

William James Earle
Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

SOME CONSEQUENCES OF SEMANTIC EXTERNALISM

by

Consuelo Preti

Adviser: Professor Jerry A. Fodor

Semantic externalism is the view that meaning and mental content are determined by relations to the world of objects and properties outside the physical boundaries of the subject of mental states. What you mean by your words - what you're thinking when you're thinking about something - is essentially constituted by the world at large. It has become customary to formulate externalism in terms of so-called twin earth cases - cases where (some kinds of) content do not supervene on inner states, but this formulation can be shown to be too limited to be of any great use in characterizing a theory of mind. A more general formulation of externalism is defended in chapter 1, one that characterizes all content.

That externalism has untoward consequences for belief-desire psychology is a familiar point, but, given the predominance of twin earth formulations, the problem of content's explanatory role is often construed as the problem of content's failing to supervene. In chapter 2 we argue that this is a mistake. Externalism in its most general

formulation has consequences for *all* content in explanation, not just content that fails to supervene on inner states. If content is externalistically individuated, then content is *redundant* in causal explanation. In chapter 2 we examine the redundancy problem and consider the options for its solution.

Part one of chapter 3 concerns the further - unnoticed - consequence of externalism that requires a reconsideration of Davidson's charge that Fregean semantic theories fail the test of 'semantic innocence.' If meaning is partly determined by reference, then what an expression refers to in an opaque context, is, ultimately, its *reference*. Part two of chapter 3 concerns the consequences of externalism for analytic truth. If meaning is partly determined by reference, a question arises as to what becomes of the classical philosophical distinction between analytic and synthetic truth; between 'truth in virtue of meaning alone' and 'truth in virtue of meaning and the world.' Chapter 3 concludes with an account of analytic truth from the perspective of semantic externalism.

ACKNOWLEDGEMENTS

I was delighted when Jerry Fodor agreed to supervise this dissertation, particularly since his ideological reservations were fairly severe. It is a testament to his integrity - and to his good-naturedness - that he was able to reconsider his opposition to various points at various stages as a result of my arguments. He was extremely generous with his time, and readily spent hours arguing with me - in person, by phone, on e-mail, in pages of point-by-point commentary. Jerry Fodor is a man who is deeply pained by philosophical positions that don't make sense to him: by forcing me to make mine make sense to him he taught me more about doing philosophy than he suspects. He was a fierce opponent and he is a good friend, and I am very grateful to him.

I would also like to thank Stephen Schiffer for the many hours we spent discussing these issues. It goes without saying that coping with his always spirited and sometimes telling criticisms helped to give this work the shape that it has.

I was greatly influenced by the many agreeable discussions I had with the following people, whose

encouragement was plentiful: Akeel Bilgrami, Malcolm Budd, Graeme Forbes, Mark Johnston, Brian McLaughlin, Gary Ostertag, and Galen Strawson.

I am very grateful to Bill Earle for his kindness, generosity, steadfastness, humor, and cleverness. Without him and his support things might have turned out quite differently. His friendship has meant a great deal to me.

My greatest debt is to Colin McGinn. The clarity, meticulousness, and faultlessness that characterize his philosophical ability has taught me more about doing philosophy than anything else. No one who has come under his influence likes to dwell on the kinds of errors, confusions and misjudgments they would have committed - and would have gone on committing - were it not for him. Certainly without his support it is doubtful this work would have turned out quite the way it has. Colin McGinn is a philosopher who is unafraid to recognize and encourage the work of others; for this and for many other things I am very grateful to him.

CONTENTS

ACKNOWLEDGEMENTS	vi
CHAPTER	
1. EXTERNALISM: FORMULATING A THEORY OF MIND	1
1. Introduction	1
2. Meaning, Reference, and Externalism	4
3. Supervenience and Externalism: A Reductio	10
4. Externalism and Twin Earth	11
5. The Limited Applicability of Twin Earth Cases	18
6. Internal Properties	23
7. Supervenience	26
8. Strong and Weak Externalism.	35
9. Weak Externalism: Presuppositions.	38
10. Weak Externalism and Exportation	42
11. Weak Externalism: Formulation	48
12. Weak Externalism and Propositional Constituency.	60
13. Externalism and Abstracta	65
14. Conclusion	69
NOTES TO CHAPTER 1	71
2. EXTERNALISM AND ITS CONSEQUENCES FOR PSYCHOLOGICAL EXPLANATION	84
1. Introduction: Causation and Explanation	84
2. The role of laws in scientific explanation.	87
3. Mental Causation	90

4. Belief-Desire Psychology	92
5. The Case of Factives: An Analogy.	94
6. Externalism and the role of content in explanation	96
7. Non-Supervenience, Extrinsicness, and Redundancy.	98
8. Non-Supervenience, Extrinsicness, and Redundancy: Entailment Relations	114
9. The Redundancy Problem: Distinctions	116
a. Syntactic properties vs. semantic properties in explanation.	116
b. Redundancy and the special sciences.	120
10. Resolving the Problem of Content's Explanatory Role: The Options.	129
a. The consistency of explanatory relevance and externalism	132
b. Syntax.	136
c. Narrow Content.	138
d. Non-Causal Criteria of Individuation.	143
e. Teleology	151
f. Rationalization vs. Causation: Radical and Moderate Views.	157
11. Conclusion.	171
NOTES TO CHAPTER 2.	174
3. EXTERNALISM AND ITS CONSEQUENCES FOR SEMANTIC INNOCENCE AND ANALYTIC TRUTH	185
1. Introduction.	185
PART ONE: SEMANTIC INNOCENCE.	189

1. Introduction:	189
2. "Two Provinces So Fundamentally Distinct": Meaning and Reference	190
3. Externalism and the Reference of That-Clauses: Exotica or the Mundane? . . .	192
4. The Varieties of Externalism: the Account Applied	199
5. Opacity, Belief and Analyticity.	201
6. Conclusion	203
 PART TWO: ANALYTIC TRUTH	205
1. Externalism and Analyticity.	205
2. Truth-Conditions and Analyticity-Conditions: A Distinction.	207
3. The Structural Complexity of Abstract Properties.	209
4. Property-revealing sense	210
5. The Account Applied.	216
6. Objections and Replies	218
a. Externalism: Proper Formulations . . .	218
b. Mentalism about Properties Opposed. . .	220
c. Externalism and Abstracta.	223
7. Conclusion	226
NOTES TO CHAPTER 3, PART ONE	229
NOTES TO CHAPTER 3, PART TWO	231
REFERENCES	235

CHAPTER 1

EXTERNALISM: FORMULATING A THEORY OF MIND

If my darling were once to decide
Not to stop at my eyes,
But to jump, like Alice, with floating skirt
into my head,

She would find no tables and chairs,
No mahogany claw-footed sideboards,
No undisturbed embers . . .

Philip Larkin, "If My Darling"

1. Introduction

In what follows, I will be formulating and defending a certain thesis about the individuation of mental states, one that claims that mental states are determined by relations between the subject of those states and the world beyond the subject.

The underlying substantive issue concerns the nature of mind. The traditional view, whose paradigm is the Cartesian model, claims that the existence and identity conditions of mental states are independent of the existence and identity conditions of the items that are the objects of those states; there is no necessary correspondence - let alone individuation relation - between states of affairs and mental states. In direct opposition to this is the view that claims that mental content is related to states of affairs

so thoroughly that, to put it vividly, a psychological subject should not be thought to instantiate certain states in the absence of the objects of those states.¹ The mind is thus no longer to be thought of an autonomous, world-independent entity.

We can call these two general views about the nature of mind *internalism* and *externalism*, and understand them to conflict in the following way (McGinn, 1989, 9):

. . . mind and world are not, according to externalism, metaphysically independent categories, sliding smoothly past each other. To regard them so is to commit oneself to an 'untenable dualism,' to marking a metaphysical boundary that does not exist. Internalism, for its part, insists upon such a duality, drawing a sharp line between mind and world. . . .

What we are concerned with in this chapter is a formulation of externalism that best captures the idea that there is a genuinely constitutive relation between the mind and the extra-mental world (suitably delineated), and just what kind of relation this might turn out to be.

Judging by the literature, at least two things might be meant by externalism²; and one formulation, we will argue, is too limited to adequately capture a robust notion of externalism. The rest of this chapter will be devoted to first, distinguishing two formulations of externalism; and second, to make good on the claim that only one of these formulations supports any general or widespread condition for content.

A brief introduction to the alternative formulations is in order. One employs the familiar device of twin earth (TE), and entails a failure of supervenience of mental states on inner states. This view takes content to be determined by a causal or contextual relation between the subject and the world outside his head, from which it follows that the mental states of two internally indistinguishable subjects fail to supervene on those internal states when the subjects inhabit different environments. The other formulation of externalism involves no entailment of non-supervenience; it is the view that only the co-presence in a world of a subject and the relevant worldly entity to which he is related is required in order to effect the individuation of his mental states.

This more general formulation of externalism is to be understood as a thesis concerning under what conditions someone may or may not possess a thought: it is generally understood to be the claim that it is a necessary condition on the possession of a thought that the subject be related to extra-cranial objects and properties - the worldly entities that are the reference of the expressions he employs.³ It follows from this, of course, that should the requisite objects and properties fail to exist, the thought in question is, in Noonan's terms (1991, 1) "not available to be thought." Note that the condition applies counterfactually as well: were the requisite objects or

properties to fail to coexist with an individual in a world, that individual would not possess the concepts held to be individuated by those entities.⁴

Now, as we will show, twin earth cases do not apply to all concepts, though of course they do apply to some; so formulations of externalism in terms of twin earth cases are too limited to be of any great use in formulating a *general* theory of mind. What will emerge is that there is more to the notion of externalism than is implied by cases where content fails to supervene.

Among the many details we will consider in subsequent sections is one so crucial it is best to mention it, even preliminarily, here. It will emerge that the distinction in formulations of externalism can be understood as turning on important distinctions concerning the notion of *supervenience*: in particular, concerning what is to be included or excluded in the supervenience base.⁵ First, however, some general remarks.

2. Meaning, Reference and Externalism

Formulations of externalism in terms of twin earth cases have occupied a prominent position in some of the literature; but there is, of course, no in principle reason to take these as *definitive* of externalism. We can begin with an examination of the intuitions behind the emergence of the notion of externalism in the literature, in order to

build up to the idea that formulations of externalism involving cases of failure of supervenience are too restrictive to adequately capture that notion.

Putnam (1975a, 215-271) was among the first⁶ to consider the challenge to traditional theories of mind now known as semantic externalism. Traditional theories are, according to Putnam, *methodologically solipsist*, a view he characterizes as follows:

. . . no mental state, properly so-called, presupposes the existence of any individual other than the subject to whom that state is ascribed.⁷

An externalist theory, by contrast, can be understood as the *denial* of methodological solipsism. Burge (1982, 117;120) clarifies this, noting that:

In some instances, an individual's having certain *de dicto* attitudes entails the existence of entities other than himself and his attitude contents . . . [Putnam] also gives examples of psychological states in the 'wide sense', and characterizes these as *entailing* the existence of other entities besides the subject of the state.

So, as the remarks by Putnam and Burge above indicate, "widely" individuated mental states, to be "properly so-called," presuppose or entail the existence of some entity other than - independent of - the subject of the state.

A few preliminary remarks are in order. Ultimately, the issue in question centers on that of *intentionality*: what our thoughts are about, captured in what our words

refer to. An externalist, as opposed to an internalist, will hold that the content/meaning of a person's thoughts/words is (necessarily) individuated by items extrinsic to her spatial contours. But stated like this, the claim is very general; too general, it might be thought. It is consistent, after all, with the claim that states of affairs in a distant galaxy are necessarily implicated in the content of one's thoughts - and consistent with that claim's being a *unique* necessary condition on the individuation of content. And what we are after, in the end, is a way of necessarily linking a subject's mental states with the material world of objects and the properties they instantiate and with which she consorts and co-exists.

And the most obvious way to proceed is by way of the reference of the expressions of the language employed by a psychological subject. Clearly what is "outside the head" will not bear on content in the relevant manner - that is, give us a way of determining how it is that content plays a role in fixing the aboutness relation - *unless* it is through the reference of mental states; these issues are naturally joined. The relation between meaning and reference is the relation between mental states and the world, given the principle that the concept expressed by a term is given by what that term means. Given this obvious point, externalism is quite reasonably conceived of as a claim about the relation between meaning and reference - and thereby about

the relation between the mind and the world.⁸

A glance at the early literature on this question bears this out: it was precisely the meaning-reference relation - and not anything to do with failure of supervenience - that preoccupied those who initially formulated the notion of externalism familiar to most.

Recall Putnam's claims in "Is Semantics Possible" (1975e, 139-152), "Meaning and Reference" (1973, 699-711)⁹ and "The Meaning of 'Meaning'" (1975a, 215-271), three articles where the notion of externalism is introduced into the literature. "Is Semantics Possible?" is an attempt to come to grips with the question "why is the theory of meaning so *hard*?" (1975e, 139):

. . . enormous progress seems to have been made in the syntactic theory of natural languages . . . Comparable progress seems not to have been made in the semantic theory of natural languages, and perhaps it is time to ask why this should be the case.

Putnam diagnoses the problem (1975a, 216-218):

Since the Middle Ages at least, writers on the theory of meaning have purported to discover an ambiguity in the ordinary concept of meaning and have introduced a pair of terms - *extension* and *intension* or *Sinn* and *Bedeutung*, or whatever - to disambiguate them . . . Suppose that there is a sense of 'meaning' in which meaning = extension, there must be another sense of 'meaning' in which the meaning of a term is not its extension but something else, say the 'concept' associated with the term. Let us call this 'something else' the *intension* of the term Something like the preceding paragraph appears in every

standard exposition of the notions 'intension' and 'extension.' But it is not at all satisfactory . . . Unclear as it is, the traditional doctrine that the notion 'meaning' possesses the extension/intension ambiguity has certain typical consequences. Most traditional philosophers thought of concepts as something *mental* . . . Secondly, the timeworn example of the two terms 'creature with a kidney' and 'creature with a heart' does show that two terms can have the same extension and yet differ in intension. But it was taken to be obvious that the reverse is impossible: two terms cannot differ in extension and have the same intension. . . .

And makes a now familiar remark (1973, 710-11):

. . . the theory that (1) words have 'intensions' which are something like concepts associated with the words by speakers; and (2) intension determines extension - cannot be true of natural-kind words like 'water' . . . [this] leaves it open, however, whether to say that 'water' in the Twin Earth dialect of English has the same *meaning* as 'water' in the Earth dialect and a different extension - thereby giving up the doctrine that 'meaning (intension) determines extension,' or to say, as we have chosen to do, that difference in extension is *ipso facto* a difference in meaning . . . thereby giving up the doctrine that meanings are concepts, or indeed, mental entities of any kind.

Those familiar with the literature will recall that Putnam was concerned to mount an attack on a Fregean semantic theory; he was not alone (cf. Kripke (1972, 253-355), Donnellan (1962, 356-379; 1966, 281-304), and Kaplan (March 1972, draft), among others). The arguments concerning the role of reference in the individuation of meaning were

crucial to that project. The implications for a theory of mind were obvious; and externalism emerged on the scene.

Now, the formulation of externalism as a claim concerning the existence-dependence of mental states on their objects - a formulation that takes its cue from the denial of methodological solipsism and which is often understood as the view that reference partly determines meaning - certainly does not entail the failure of content to supervene on an individual's internal states; this view does not entail TE thought experiments. The TE formulation has occupied a prominent position in the literature; so prominent that some take it to be *essential* to formulations of externalism. Our preliminary remarks suggest that this is a mistake: and, in fact, there is no reason to accept the TE formulation as definitive of externalism, as we shall next argue.

It is worth noting that, since much of the notice in the literature regarding externalism as a theory of content devolves on its repercussions for content in belief-desire explanation, it is all the more pressing to clearly establish that the orthodox understanding of externalism, formulated in terms of content's failure to supervene, is, at best, limited. This, of course, suggests that the problem of content in explanation is not - exclusively - the problem of the failure of content to supervene on inner states. I will show (see chapter 2) that the more general formulation

of externalism likewise results in explanatory role problems for content, one of three consequences for this view that I examine (see chapter 3). In sum, I contend that the true significance of externalism as a theory of mind might be better appreciated if it can be shown to be true of, and to have repercussions for, content in general - something that the influence of twin earth may have obscured.

3. Supervenience and externalism: a reductio

We can begin with a reductio of the idea that there is an essential connection between failure of supervenience and the notion of externalism, as a preliminary salvo. The Cartesian theory of mind has it that mental substance and physical substance are utterly distinct - in fact, mental states can be what they are independent of there being bodies (heads with brains in them, say) at all. On this view, it is perfectly conceivable - Descartes even thought it was *desirable* - that one's mental states be what they are whether or not the objects of those states exist. Now, this is a view that entails, quite dramatically, a failure of the supervenience of the mental on inner states; but it is really a gross and unwarranted perversion of the Cartesian position to claim that from this failure of supervenience follows externalism about content. Descartes is, notoriously, an *internalist*: what states one is in are not essentially determined by any relation to the world.

Of course, it might be thought that Descartes is an externalist with respect to God-thoughts, for these have the their content in virtue of God, the only being of sufficient magnitude to cause those ideas in us - and God is an extra-cranial entity. (someone who clings to the TE formulation of externalism will not see this, since what you're thinking when you think about God won't vary depending on what world you're in; more on this below). This would be confused, however: the relevant thing to remember is Descartes's *mentalism*. Ideas in the mind - whose origin is innate - are the objects of thought and knowledge. That these Ideas happen to correspond to the world (thanks to the non-deceptive nature of God) is not to say that these Ideas are vulnerable to fluctuations of any sort in the world. God's job is to ensure veridicality, and hence certainty; the *individuation* of content is not His concern.

4. Externalism and Twin Earth

An unclouded consideration of the essence of TE cases and what, if anything, they can be taken to establish with respect to externalism as a general theory of mind, is next in order.

Putnam (1975a, 214-271) uses TE cases as an illustrative device in an argument to support the claim that what a person's utterances mean cannot be a function of his internal properties alone: a claim, as we have said above,

quite plausibly understood as the denial of methodological solipsism. (Nowhere, of course, is the TE formulation of the issue offered by Putnam as a *definition* of externalism). Surrounding a psychological subject are objects (and, *n.b.*, the properties they instantiate); these are what he refers to by his words, what his thoughts are about. The question is: is it consistent to suppose that two subjects who share all the same internal properties share the same conceptual ones?

Putnam makes the point - aided by TE cases - that it is not consistent: it is stretching plausibility to maintain that one's words refer to anything - that one's thoughts are about anything - but the items in the extra-cranial environment with which one is in (causal) contact. The essence of a canonical TE case is thus to hold the internal states of an individual fixed (or one individual at different times, or entire communities of individuals) and to vary the environment. The upshot is that a variation in the environment results in a variation of a mental state.¹⁰

Burge (1979, 73-121) makes an amendment to the basic sketch.¹¹ Among the items in one's extra-cranial environment are other people, who perform a range of linguistic acts. Isn't it plausible to suppose that the meaning of one's words and the content of one's thoughts might also be subject to variation in the linguistic

practices and conventions in the social environment? Thus, consider two individuals who inhabit two communities where the linguistic practices relevant to the use of the term 'cramps' differ. In community 1, 'cramps' correctly applies to a uterine complaint alone. In community 2, 'cramps' correctly applies to a uterine as well as a general muscular complaint. Now consider two male individuals - identical in every internal property - who utter the sentence "I've got the worst cramps again, doctor." In community 1, the doctor would presumably correct the speaker, saying that whatever is wrong with him it *can't* be cramps. In community 2, however, his doctor would simply prescribe the necessary sedative.

We can critically examine number of things about the basic idea, which may be useful in getting a clearer grasp on the deeper intuition for which these cases are meant to be an argument. First we will consider the nature of the relation between Putnam-type TE cases and Burge-type cases (which often are themselves formulated using the TE device), and then move on to show that there is content for which it is difficult to maintain that TE cases can be erected.

First: Putnam-type TE cases can be understood in terms of the appearance/reality distinction (more on this below). Both twins are in the presence of two substances that present the very same *appearance*, but that have - as it is sometimes called - different *essences*.¹² These essences

are in principle discoverable by the experts, to which we may need to defer for the sake of complete accuracy. Burge-type cases, however, at least according to Burge, are to be understood in terms of an individual's incomplete grasp of the meanings of his own words (1979, 79):

The argument can get under way in any case where it is intuitively possible to attribute a mental state or event whose content involves a notion that the subject incompletely understands. As will become clear, this possibility is the key to the thought experiment.

Questions to do with distinguishing some *cubic zirconium* from some *diamond*, or a glass of Clorox from a glass of water, can be settled - so, likewise, can questions to do with what the words that refer to those entities mean. For settling questions to do with what the meanings of words like 'arthritis' or 'cramps' are, different standards - *community standards* - are thought to apply. Vary the community standard and you vary the meaning of the word.

Now, one thing to notice is that Burge's claims about the role of the social environment in the individuation of a mental state are not such as to impose a *necessary* condition on fixing content, although this is, more often than not, misunderstood.¹³ This strikes me as an important disanalogy with respect to Putnam's conclusions about the role of the extra-cranial environment in fixing the nature of mental states.

Putnam's TE examples are, strictly speaking, best

interpreted as making a necessary claim concerning *distinctions* in content-possession: twins whose environment is not the same will (necessarily) not share the same concepts. But it is also plausible to suppose that Putnam's claims about the relation between the environment and the mind are to be understood as imposing a necessary condition on concept-possession itself.¹⁴ Certainly the formulation of the view as the denial of methodological solipsism quite explicitly exploits an entailment relation between the existence of a head-independent, environmental entity and the mental state that takes that entity as an object. TE cases are not so explicit, but, as many would agree, it is perverse to take them otherwise.

The key point is this: Putnam's formulations of the notion of externalism are quite plausibly understood as invoking a necessary connection between one's mental states and the environment one is in. No such necessary connection is present in the Burge-type cases, however. For one thing, social externalism (as we might call Burge's view) is at the very least logically independent of environmental externalism (Putnam's view), so considerations may in principle be made with respect to one that do not impinge on the other.

Consider that there is no claim in Burge to the effect that one *could not have* the concept 'arthritis' in the absence of a social environment. As he himself makes clear,

his thought-examples turn on *defective* possession of a concept - it is members of a community with incomplete understanding of a given concept whose mental states are, as it were, vulnerable to variations from the social environment. Nothing in Burge's claims prevents someone in possession of the full definition of 'arthritis' to genuinely so possess it quite independent of any of his peers.

There is another, more important point to notice. From the fact that meaning is subject to variations in an individual's *social* environment it does not follow that meaning isn't "in the head." The fact is that what individuates the meaning of a particular individual's expressions may just as well be *in the heads* of one or more of those who comprise his social environment - a clear case of anti-individualism¹⁵ but internalism about meaning. Burge never clarifies this point, appearing to think that the roles individuals other than oneself may play in the fixing of the content of one's mental states is sufficient to promote externalism in the relevant sense. It is hardly the case, however, that other minds are sufficiently externalistic - merely by being *other* - to add the other dimension to externalist criteria of content individuation that Burge seeks. It is hard to see that merely being not in *my* head but in the head of another is going to capture the intuition at work behind the notion of externalism.

Further, and conversely, from the thesis that meaning is subject and sensitive to variations in an individual's environment need follow nothing about the role of the individual's social environment in the determination of meaning. So care must be taken here when applying these cases to establish externalist claims.

Now: is it the case that the Putnam/Burge TE cases apply to *all* concepts? They are taken to, at least by Fodor (1988, 27):

. . . I'll assume that the Burge story shows that if the Putnam story raises any problems for the notion of content, then the problems that it raises are completely general and affect all content-bearing mental states.

Putnam seems to concur (1975a, 242):

So far we have only used natural-kind words as examples; but the points we have made apply to many other kinds of words as well.

Burge himself, on the other hand, appears to have realized that Putnam-type cases are of limited applicability, noting (1979, 118):

The argument regarding the notion of water that I extracted from Putnam's paper is narrower in scope than our argument. The Putnam-derived argument seems to work only for natural kind terms and close relatives.

However, he seems to think that his own cases apply quite generally (1982, 117):

The Twin Earth thought experiment may work only for certain propositional attitudes. Certainly its clearest

applications are to those whose contents involve non-theoretical natural kind notions. But the arguments of 'Individualism and the Mental' suggest that virtually no propositional attitudes can be explicated in individualistic terms.

There is reason to think, however, that this general applicability thesis isn't true. First: since the Putnam cases don't entail the Burge cases, nor the other way around, one would have to find some way to defend the view that the individuation thesis supported by these cases is completely general. Second: even if Burge-cases can be applied to a wider range of cases than those of Putnam, it far from follows that Burge-type cases apply to most, let alone all concepts. Recall that our concern here is to formulate a notion of externalism that applies to as many concepts as possible. It is my contention that the kind of formulation of externalism exemplified by the Putnam and Burge cases is too limited to be of any great interest for issues concerning mind and mental content. An examination of a few concepts for which TE cases can't be erected is next.

5. The Limited Applicability of Twin Earth Cases

Putnam's well-known summary of his view is this (1975a, 227): "Cut the pie any way you like, 'meanings' just ain't in the head!" McGinn makes a laconic comment (1989, 31):

In fact this conclusion exaggerates what has been established, since only some

meanings have been argued to be environmentally determined, not *all*. The appropriate conclusion should have been that a proper subset of meanings are not in the head - a somewhat less resounding announcement.

Let us see whether this is so: is there content for which for which Putnam-type TE cases cannot be erected? If Putnam-type TE cases can't be erected for some concepts, do Burge-type cases take up the slack? If not, are there concepts for which externalist individuation does not - or even cannot - apply? We will answer these questions respectively: yes, no, and no.

TE cases, remember, are cases where internal facts are held constant and the extra-cranial "surround"¹⁶ is varied. So what can we make of cases where the relevant extra-cranial entity can't be varied? McGinn gives a fairly comprehensive list (1989, 47-100): that which is referred to by complex concepts, formal concepts, psychological concepts (concepts of mental states), and perceptual concepts. To this we could add also the ethical and aesthetic environment.

The most obvious instance of non-conformity to a standard TE case concerns formal entities like number. How is it possible to erect a situation where the mathematical environment is varied but the internal factors - say the way that environment appears to those on Earth and on Twin Earth - stay the same? It seems practically incoherent to say that 'prime number' or the logical expression 'if and only if'

refer to different things on Earth and Twin Earth but that both communities have the same qualitative experiences with respect to those entities.

What about, analogously, aesthetic concepts such as 'beauty'? Does it make any sense to claim that on Earth and Twin Earth the aesthetic environment is (really) different but *appears* the same to both sets of people, so that the word 'beautiful' refers to really quite different things (maybe ugly things) that don't seem at all different? Could the *moral* environment really be varied while appearing the same both to me and my twin, so that when she applauds someone's integrity, she's doing something I might in fact recoil from?¹⁷

Davies (1992, draft, p.3) makes a few remarks concerning the applicability of what he calls modal externalism (what characterizes concepts that are subject to TE cases) that are directly relevant here:

. . . perhaps we should consider . . . ways in which it might turn out to be impossible to generate the 'Twin Earth' examples that would establish modal externalism. One kind of case would be where there is a necessary connection between the relevant features of the environment E and X's inner constitution, so that a situation with environment E' instead of E is inevitably a situation in which there is no duplicate of X.

If all this is correct, then not only do TE formulations tend to come apart under scrutiny - even for those concepts for which they are generally thought to apply

- but, more importantly, there are any number of concepts for which the preconditions for setting up such cases are not met. There is, as a result, no reason to suppose that TE cases - at least Putnam-type TE cases - do apply generally.

Given the logical independence of Putnam- and Burge-type TE cases there is no reason to think that what the Putnam cases don't cover the Burge cases will, but it is worth pointing out that social externalism doesn't apply to all concepts either.

Burge points out (1979, 79) that his cases depend on the individual's only partially understanding the term that she employs - but of course, Burge-type cases are slightly more complicated than that. Burge never says that it is *sufficient* to get one of his cases off the ground that one be deficient in a particular concept that one possesses or employs. What is actually operative in these cases is that the deficiency be such as to allow for cases where I and my counterpart *both* have a concept - but not the same one. My counterpart and I both have a concept we express with the word 'arthritis' - but given that, in the counterpart community, 'arthritis' is used to refer to an ailment of the joints *and* muscles, rather than the joints alone, we do not have the *same* concept, in spite of our internal identity.

Now, imagine running a Burge case for a color concept - say 'purple.' Could it plausibly be the case that a person employs a color concept only partially correctly:

say picking out purple things correctly six out of ten times, and picking out yellow ones the other four times; and that the relevant community standards allow 'purple' in fact to apply to both purple and yellow objects? This just can't be right: a person who so grasps the concept 'purple' must be said not to grasp it at all.¹⁸

Burge cases are also unlikely to go through for psychological concepts: for concepts of mental states. For consider: we can stipulate the counterfactual community as having different experiences from mine that they describe with the same words. The words 'experience of red' apply to, on *Earth*, an experience of red, but on *Twin Earth*, they apply to an experience of green. The same words express different (experiential) concepts. But now notice that if the counterfactual community can fix my counterpart's experiential concepts - the way it's supposed to work for others - then her self-ascriptions with respect to those concepts will be *false*. Her concept 'experience of red' is fixed by the counterfactual community, where it means 'experience of green' - because that is the kind of experience the community is having on *Twin Earth*. But the experience that she's having is a red experience - since that's the experience *I'm* having and we are *internal duplicates*. The same points apply vice versa for me, obviously. My concept 'experience of red' is - supposedly - fixed by *my* community, according to the experiences that

they are having - and suppose these are experiences of *blue*. My twin's community, however, has experiences of *red* that they describe with the words 'experience of red.' I have the *same* experiences as my twin - since we are internal duplicates - so when I ascribe to myself the experiential mental states under consideration here *I am wrong* - since the words 'experience of red' on Earth apply to experiences of blue. Surely it is vastly implausible to think that anyone is so hyperfallible about the content of one's *own* experiences - and therefore about the concepts used in referring to those experiences. Likewise, it is implausible that Burge cases go through for such concepts.¹⁹

There is another aspect of externalism that needs to be clarified, as it has important repercussions for the distinction in formulations of externalism that we shall be exploiting in later remarks. The issue here concerns the understanding of supervenience on internal properties or states: how is the idea that content - in some cases - fails to supervene on such states most accurately to be understood? In the next section (section (6)), we will make some remarks concerning the best understanding of *internal* properties. In section (7) we will reconsider the notion of *supervenience*.

6. Internal properties

We can begin by taking note of a few interesting

details. Burge (1979, 79), for instance, gives an inventory of 'internal' factors that he takes as relevant to establishing externalist content individuation:

internal qualitative experiences,
physiological states and events,
behaviorally described stimuli and
responses, dispositions to behave, and
whatever sequences of states [mediate]
. . . input and output.

In order to assess the true contribution of TE cases in support of externalism as a general theory of mind, however, it is important to characterize 'internal' factors in such a way as to avoid begging the question. It won't, for instance, follow from the fact that a state is inside the head that it is thereby 'internal' in the relevant sense. McGinn (1989, 2) notes:

it is important to notice that
'internal' is being used as a term of
art . . .; internalists, in the intended
sense, do not literally locate the mind
inside the head (though they may do);
they assert, rather, that mental states
are determined by facts relating to the
subject considered in isolation from his
environment - by facts about *him* . . .
internalism is best seen negatively as
the denial of externalism; it is the
role of the environment in fixing the
nature of mind that is centrally at
issue.²⁰

Now, it is natural to take the base domain in issues regarding the fixing of content as physiological - even neurophysiological - but we ought to be pedantic. There are at least three kinds of things that could plausibly count as 'inner' or 'internal' in the sense needed to drive the

intuitions about content illustrated in the TE cases. These are: 1) internal states of the body and brain 2) behavioral dispositions 3) local stimulations at the sensory receptors. And, what's more, it is certainly *open* to someone to claim that all three are individuated *relationally*. What's relevantly 'internal' in these discussions should not be thought to be fixed when the *geography* is fixed.

It seems that inaccurate formulations of the upshot of the TE story can lead to pitfalls concerning grasp of the more general nature of externalism that I shall shortly be urging. Consider Stich (1978, 575, emphasis his):

. . . the principle of autonomy states that the properties and relations to be invoked in an explanatory psychological theory must be supervenient upon the *current, internal, physical* properties and relations of organisms.

The kind of failure of supervenience thought to be illustrated by TE cases won't necessarily follow even if the states on which content is supposed not to supervene are thought to be neurophysiological. Again, that such states are internal to the organism - in a geographical sense - isn't enough to *guarantee* that they are *non-relational*; and it is *this* sense of 'inner' or 'internal' that is required to get failure of supervenience.²¹

Further: surely it is at least logically possible that content should supervene on *qualitative* states, which a) needn't themselves be thought to supervene on the physiological (disembodied minds, etc.); or b) might well,

for all we know, be themselves non-relationally individuated.

It is interesting to note that in the usual formulations of the TE thought experiment consideration as to whether qualitative states really are internal in the relevant sense is rarely raised; these kinds of states are generally assumed to be paradigms of the internal (Cf. Burge's inventory above). In fact, a short summary of Putnam's conclusions with respect to twins is that although both share the same phenomenological states ("water is a tasteless, colorless, liquid," they say) they do not share the same contentful states as expressed by that sentence. But this does beg the question: what if *phenomenological* states are externally individuated? It takes argument to show that they are not, if they are not; this cannot simply be presupposed.²²

6. Supervenience

The canonical formulation of externalism, as we know, has it that content fails to supervene on inner states. What we will be formulating and defending shortly is that there is another - much more general - formulation of externalism that *denies* failure of supervenience. On this more general formulation of externalism, content is - consistently - *both* externalistically individuated *and* supervenient.

It turns out, in fact, that the notion of

supervenience is ambiguous enough to lead to some genuine misunderstandings about the general notion of externalism. What we shall do in this section is distinguish between three notions of supervenience, as to avoid even *prima facie* inconsistencies. It goes without saying that in the subsequent discussion of *weak* externalism, the distinction between the three notions of supervenience must be kept firmly in mind.²³

First things first: obviously, no claim about supervenience or non-supervenience will in itself involve any claims about *externalism*. Supervenience is a metaphysical relation between two domains: it is the claim that one domain supervenes on another if there can be no difference in the supervening domain without difference in the base domain, and can be brought to bear in the analysis of almost anything.²⁴ When, however, this notion is brought to bear on issues concerning the *aboutness* relation, then the question is as follows: what facts are such that when you fix them, the aboutness relation holds? or, what does the aboutness relation *supervene on*?

This question ultimately concerns the interaction of three things: the internal states of heads, the aboutness relation, and the objects out there in the world. It is sometimes put by asking whether fixing the internal states of the head is sufficient to guarantee the aboutness relation under certain conditions. What tends to be

presupposed - completely plausibly, of course - is that the object that the mental state in question is about exists. An analogy helps with this.

Consider the 'loves-relation.' Suppose our question is whether the 'loves-relation' between *x* and *y* supervenes on the intrinsic states of, say, *x*. What we are obviously asking here is what makes it the case that *x* loves *y* and not *z*, and it ought to go without saying that we are presupposing that *x* and *y* (and even *z*) exist. Now we might of course have been interested in a more abstract question, one having to do with the loves-relation and its necessary and sufficient conditions, and that kind of inquiry would not necessarily presuppose the existence of any beings that instantiate the relation. But our previous question was not that one: it was the question what is it about *x* (if anything) that is sufficient to guarantee his loving *y* and not *z*.

Now, externalism as a general claim can be understood, as we noted in section (1), as the view that a mental state essentially involves being in a relation to the object of that state - that mental states are *relational*.²⁵ In other words, given that an object exists, what we want to know is what it is about a subject (if anything) that makes her thought about *a* and not about *b*. Is it sufficient that she be in a particular physiological state *s* - that is, does her thought being about *a* rather than *b* supervene on her

intrinsic states? Or is there more to the story? Could the causal/environmental relations that she bears to a play a relevant role? What about these kinds of relations themselves? Are they conditions that must met in order to claim that a mental state is relational? Or can a mental state be relational without them?

These remarks set up the context against which it is imperative that the distinction to come in formulations of externalism be understood. Canonical TE cases make the point that whether you're thinking about *water* as opposed to something else does not supervene on your intrinsic states alone. Intrinsic states are not sufficient given that your twin has just those states and yet is not thinking about water; how could she be? She bears causal/environmental relations to a certain substance that is not water but something else. And it is these causal relations that have to be added, in TE-type cases, to reach an adequate supervenience base (i.e. sufficient condition).

Now, the analogy with the question concerning the 'loves-relation' specified above makes it absolutely clear what the issue concerning TE cases is all about. The question there concerns the issue whether, *given* the existence of XYZ and H₂O, whether you're thinking about one or the other is what fails to supervene on inner states. The intrinsic states, familiarly, are not sufficient: what is further necessary is certain causal/environmental relations

that will select (as it were) one thing rather than the other.

What we want to defend is a notion of externalism that goes (well) beyond TE cases, since, as we saw above, these kinds of cases cannot be erected for many kinds of content - that is, no causal dependence or environmental theory will be correct or apply to them. And a very natural way to put this will be to say that many mental states are *relational* without being causally/environmentally relational. In fact, we can go further: we will show that there is a sense of supervenience that is perfectly consistent with the claim that content supervenes but is nevertheless relational.

Now, on the face of it, there seems to be a glaring problem. Externalism is the claim that mental states are relational: it is a necessary condition on being in a mental state *m* that the object of the state exist. But to say that mental states *supervene* on intrinsic states means - in one sense of supervenience - that intrinsic states are *sufficient* to fix what mental state you're in. So how can a mental state be both supervenient and externalist?

The answer is that the notion of supervenience at work throughout this issue is ambiguous enough to support three different readings. In the formulation that follows of weak externalism, the notion of supervenience at work there could be understood, roughly speaking, as a kind of shorthand. What we need to do is specify what the supervenience base

includes or does not include in order to make certain crucial distinctions between states that supervene, states that don't, and to formulate a notion of externalism that straddles that distinction. Once the supervenience base is specified for a particular case, all we need to do is to remember what the term 'supervenience' for that case is meant to include or exclude. The same term will be used, but it will mean slightly different - but importantly different - things.

We distinguish between three cases:

1. *Pain states*: when we say that pain states supervene on physiological states, we stipulate that the supervenience base consists solely of the physiological states. It is sufficient, to fix the pain state, to fix the physiology: there are no other necessary conditions.

2. *Canonical TE cases*: Here, we must understand the supervenience base to consist of three things: the intrinsic facts about the subject; the facts about the world - the existence of the objects of the relevant mental states - and, further, the causal/environmental relations that the subject bears to the objects of her mental states. These three are necessary for determining whether someone is thinking about one thing rather than another; and they are *jointly* sufficient. Content (in some cases) fails to supervene on intrinsic states in the sense that intrinsic facts alone are not sufficient to fix the aboutness

relation; a causal/environmental relation is also required.

3. *The intermediate case: weak externalism.* There is a position in logical space that denies that causal/environmental relations are part of the supervenience base. The claim is, roughly speaking, that mental states can be relational (externalist) without being causally/environmentally relational. What we need to stipulate for this kind of intermediate case is that the supervenience base will include intrinsic facts, and it will include facts about the world - the existence of the objects of the mental states in question - but it will exclude any causal/environmental relational facts. The upshot is that there is room for the idea that being in the very same intrinsic physiological states is sufficient to fix mental states, without ruling out the possibility that the states are relations to - but not causal relations to - head-independent entities. That is, no further property of the object needs to be instantiated, in contrast to what holds for content subject to TE cases.

Keeping the threefold distinction above firmly in mind puts us in a position to amplify - and refute - the objection alluded to above. Briefly, one way of understanding the issues here might proceed as follows: since, according to externalism, aboutness is relational (it necessitates there being an object of the intentional mental state), then it is quite false to suppose that aboutness

supervenies. The latter claim means that intrinsic states are sufficient to fix the mental state in question, but this is inconsistent with the claim that the mental state is relational. To say that a mental state supervenes on inner states is to say that the state is *not* world-involving; the only necessary condition for fixing that state relates to the head. How can it make any sense to say that a mental state is both supervenient and world-involving? It might be possible to make those two claims consistent by claiming that intrinsic states somehow guarantee the existence of the object of the state, but this is, of course, absurd.

The reply to this (admittedly rather perverse) reading of the question concerning the more general formulation of externalism is now obvious, given our distinctions above. The objection presupposes an understanding of supervenience such that the only things that are included in the supervenience base are intrinsic states. But there are two other possible formulations of the supervenience base: one, the familiar TE case, where there are *three* necessary and jointly sufficient conditions included in the supervenience base; and the other, middle of the road position, which includes not only intrinsic facts in the supervenience base, but also the existence of the *objects* of the mental state. What is not included in this latter position is, of course, the causal/environmental relations to the objects in question; and it is just here that the two views can be

essentially distinguished. And it is *this* distinction - no other - that is the focus of our discussion.

Of course, it should be fairly obvious that the entire issue concerning the aboutness relation and what it supervenes on presupposes that the objects of the intentional state exist: after all, intuitively, if they didn't, there wouldn't be such an instantiated relation. The position here is exactly on analogy with the remarks made earlier concerning the 'loves-relation.' Obviously, if what we are interested in is what facts about that relation (as it is borne by *x* and *y*) supervene on what facts about *x* or *y*, then it quite obvious that we are presupposing that both individuals exist: what sense would it make to ask the question "what makes her love *him* rather than the other guy?" if there wasn't anyone you were talking about?

As we said above concerning a familiar TE case, the issue is not to do with the *existence* of H₂O or XYZ somehow failing to supervene on intrinsic states, rather, the question concerns *which* of those substances you can be said to be thinking about under certain conditions. What we have shown here is that there is logical room to hold - given a specification of supervenience that builds in the perfectly obvious presupposition that the object of the mental state exists - that internal facts are sufficient (in conjunction with the object) to fix the mental state in question; but the mental state is no less relational for all that. Again,

the only facts about *the subject* that are part of the supervenience base are *intrinsic* - no extrinsic causal/environmental relations are required, in contrast with the lesson of the TE case.

8. Strong and Weak Externalism

The question to consider now in greater detail is whether the fact that TE cases - both the social and environmental varieties - don't apply to all concepts means that externalism need be abandoned as a general theory of mind. The answer, as we have suggested, is no.

We can best think of externalism as a genus of which there are two species. Following McGinn (1989), we can distinguish externalist claims into two sorts. *Strong externalism* is to be understood as the species of externalism that entails non-supervenience. *Weak externalism* is to be understood as the species of externalism that claims that the (necessary) existence and identity conditions of content are those of the objects of the mental state with that content. The key difference is that strong externalism claims that the *local* environment inhabited by a subject is what determines her mental states; it is this *part* of the environment, after all, with which she has contextual and causal contact; what (some of) her thoughts are directed onto, what (some of) her words refer to. Weak externalism, however, claims that the (necessary) relation

between mental states and the environment is one between the subject and those parts of the head-independent environment to which her words refer, not just those with which she *locally* consorts.

The difference between strong and weak externalism ought not to be thought of in terms of the degree of externality that applies to concepts, but should rather be thought of as a distinction that concerns the conditions under which content - externally determined content - gets fixed. McGinn notes (1989, 8):

The distinction between the two sorts of externalism arises out of the simple point that something can exist or be instantiated without existing or being instantiated in a given subject's environment, where the notion of environment is taken causally or contextually, thus making people on twin earth occupy a different environment from us. Co-existence in a possible world is not the same as co-presence in an environmental niche. These two conditions coalesce only for the case of God: because of his omniscience and omnipresence the whole of reality is his environment - which is why you cannot run a twin earth case for God.

The best way to understand the difference between strong and weak externalism is as follows. TE cases (and the concepts subject to them) are meant to counter the Fregean picture that what one means by one's words is to do with, exclusively, facts about oneself. The anti-Fregean picture - externalism - claims that mental states are determined by relations to entities that are independent of the subject of

those states. Twins make the case that being identical in every local (non-relational) respect does not always guarantee sameness of mental content. Content doesn't (always) supervene on inner states, since we can stipulate plausible cases where all the inner states are held stable but a subject is quite obviously not in a position to be ascribed the contentful state.

Now, I can alter the natural kind environment (at least in principle) - so taking natural kind concepts with me - but no such alteration is even in principle possible with respect to, for instance, the ethical, formal, or aesthetic environment. So the respective concepts will supervene on inner states. They can't be pried apart: it isn't coherent to argue that what my twin means by 'not' isn't what *I* mean by 'not' because on TE 'not' is really *F* . . . although it looks like/feels like what it does to me here on Earth (*G*, say).

The upshot is that my twin and I are in the same mental states when it comes to 'not,' 'beautiful,' 'two' and the like. Fix the inner non-relational states and you fix the mental states when it comes to these kinds of concepts. Anyone who takes externalism to be equivalent to failures of supervenience will of course deny that externalism applies to these concepts: these don't fail to supervene. But this is, of course, nonsense.

Externalism is a quite general individuation thesis,

like many others. The general claim is that content is essentially determined by relations to non-mental entities that are independent of the subject of the mental states in question. And what could be more obvious than that being in a mental state whose object is the number two or the property 'beautiful' is to be in a relation to head-independent entities? The objects of aesthetic, mathematical and ethical beliefs are *properties*. The most sensible view to take about properties is realism and non-mentalism with respect to them.²⁶ So what you're thinking about when you're thinking about 'not' is determined by something that is not in your head: a perfectly straightforward - and recognizable - formulation of externalism.

Given our remarks about the general applicability of TE cases, we now are in a position to appreciate that for content for which TE cases can be erected - strongly external content - weak externalism will also be true. But not all content for which weak externalism is true is content for which strong externalism will likewise be true. We now turn to arguments in support of the general applicability of weak externalism.

9. Weak Externalism: Presuppositions

There are at least two ways that weak externalism might be understood. In this section, we will consider a semantic argument for weak externalism; in the next, a

metaphysical one. Both are, as it might be surmised, essentially two sides of the same coin.²⁷

Weak externalism can be derived from some straightforward - although not uncontroversial and not uncontested - semantic principles.²⁸ By way of scotching at least preliminary misunderstanding, let me say that it is imperative to understand at the outset that weak externalism should be understood to result from these principles taken together; each one in itself is not sufficient to establish the claim. The semantic argument can be most simply understood by way of the following three presuppositions:

(i) Externalism is the claim that there is an existence-dependence between mental states and the objects of those states - a characterization of mental states that, as we have seen, is consistent with failure of supervenience formulations, but that also applies to mental states that do not fail to supervene.

(ii) A Russellian semantic background, extended, however, to general thoughts. According to Russell, singular terms are meaningful on condition that they refer; this condition applies to predicates as well. It goes without saying that also presupposed here is an ordinary materialist metaphysics, wherein singular terms refer to ordinary worldly particulars, and predicates to the properties those particulars instantiate. Quantifiers are to be interpreted here as objectual rather than substitutional.

(iii) Exportation - inferring the relational form of a belief sentence from the notional form - is to be considered a legitimate inference.

The first of these presuppositions is definitional, and we have said a great deal about it already.²⁹ We turn to the second.

It is best to start with a rough sketch of the line of argument. First, a brief reminder of Russell's claims regarding denoting expressions, as summarized by Strawson (1956, 222):

Of logically proper names Russell says or implies the following things:

- (1) That they and they alone can occur as subjects of sentences which are genuinely of the subject-predicate form.
- (2) That an expression intended to be a logically proper name is *meaningless* unless there is some single object for which it stands: for the *meaning* of such an expression just is the individual object which the expression designates. To be a name at all, therefore, it must designate something.

Now, presupposing the Russellian characterization of genuinely denoting expressions summarized by Strawson above,³⁰ consider first the truth-conditions of

- (1) That dog is a dachshund

These are such that (1) is true when the referent of 'that

dog' satisfies the predicate 'is a dachshund' and false when it doesn't. Should a referent for the expression 'that dog' fail to obtain, (1) fails to express a proposition.³¹

Now, we can, quite uncontroversially, symbolize (1) in standard predicate calculus notation as

(2) Fa

to which we can further apply existential generalization, a valid rule of inference in first-order predicate logic, which yields

(3) $(\exists x)Fx$

(2), that is, has what are known as existential entailments.

But now consider the behavior of a demonstrative expression inside a that-clause. The question arises: does it retain its existential implications when embedded? For existential generalization to apply just as readily to the embedded demonstrative in

(4) Keith believes that dog is a dachshund

would be for it to follow that there is something in particular such that Keith believes that thing to be an F . And given that there is an existential entailment between

(1) (symbolized as (2)) and (3), then, intuitively, there is - on grounds of uniformity, if nothing else - an analogous case to be made for the *retention* of such ordinary existential entailments of expressions in opaque contexts. So untutored intuition might suppose. And notice that, if such an inference were legitimate, this would amount to - not forgetting our other presuppositions, of course - a capsule statement of weak externalism.

For consider that if a sentence *p* is true, then its implications have to be true. So - presupposing the legitimacy of inferring the relational from the notional form of a belief sentence - it is clear that if the belief-sentence *p* is true, then its existential entailments are also true. Add to this the Russellian theory of denoting expressions, and belief-sentences can be characterized as existence-dependent on their objects. Were no such objects to exist, then no such belief sentence could be ascribed, since no proposition would be expressed.³²

The lie of the land ought now to be more or less clear. We move on to details.

10. Weak Externalism and Exportation

The third presupposition in our formulation of weak externalism makes use of the notion known as exportation. Exportation is, of course, a thorny issue, details of which we consider below. In spite of the complications³³ -

which go beyond the scope of my project to consider, let alone adjudicate - the key idea is quite intuitive. We ought to keep firmly in mind that it is certainly an option, logically speaking, to insist that exportation is a valid inference. This intuition is consistent with a *Russellian* approach to the issues, as opposed, of course, to Fregean accounts - hence my earlier emphasis on the *combination* of the presuppositions in the formulation of weak externalism.

Some remarks about exportation are in order. Quine noticed (1975b) that in ordinary thinking about propositional attitude idioms like belief, it was appropriate to distinguish between what he called *relational* and *notional* readings of ambiguous sentences like

(5) Keith believes someone is an artist.

The *relational* (sometimes called *de re*) reading of (5) emphasizes the subject's relation to an object and is rendered with the existential quantifier outside the scope of the propositional attitude verb³⁴:

(6) There is someone believed by Keith to be an artist.

The *notional* (*de dicto*) reading of (5) captures instead the subject's relation to a proposition, and is rendered with the quantifier inside the scope of the

propositional attitude verb:

(7) Keith believes that there exist artists.

Now, Quine comments, *inter alia*, on a certain inferential move which is intuitively correct and appears reasonable, at least in some cases, but which unfortunately creates a theoretical headache. It appears difficult to get an adequate semantics for the inference from the notional form of a sentence like (5) to the relational form; the inference known as exportation.

This can be understood with the aid of a few intuitive points concerning truth-conditions. As we said above, the truth-conditions of (1) are such that the reference of the singular term 'a' must satisfy the predicate 'F.' This makes it legitimate to existentially generalize on the singular term position: what is being expressed by (1) is that what is referred to by 'a' satisfies 'F'; so, we can infer that there is an x such that Fx.

Things become less straightforward, however, when (1) is embedded in a propositional attitude context. For, as is well-known, the truth-conditions of the embedded sentence would appear now *not* to be the same: if they were, we could unhesitatingly substitute a co-denoting term 'b' for 'a' and expect that the truth-value of the whole sentence would remain unchanged. But of course, the stability of truth-

value is not guaranteed.

The truth-conditions of the unembedded sentence are made in terms of the object and whether it satisfies the predicate: no mention whatever is made of the way in which it is presented or referred to. But in an opaque context, the way in which an object is referred to makes a difference to the truth-value of the belief sentence: it is the very proposition believed that is relevant to the truth-value of the whole sentence.

These are familiar points concerning the opacity of contexts such as belief. But now the difficulty concerning quantifying into these contexts emerges. When from (7) we infer (6) it seems we are in effect proceeding as if the truth-conditions of an opaque occurrence of singular term are just what they are when the occurrence is transparent.

For what is being claimed when existential generalization is applied to (1) is that there exists an object and it satisfies the predicate. All that matters is that the object exist and that it satisfy the predicate - the way in which it is referred to does not make a difference. This suggests, in the interest of uniformity, that existentially generalizing on an opaque occurrence of a singular term is to treat that occurrence as if all that matters is whether the object exists and whether it satisfies the predicate. But failure of substitutivity of co-denoting expressions shows that the way in which the

object is referred to *is* relevant. Hence the puzzle about quantifying in.

This, roughly speaking, acknowledges the controversy: as I have said, I will not attempt to adjudicate it, let alone weigh in on any side, although a few remarks are in order.³⁵ What is at work here is the intuitive idea that expressions do have straightforward existential entailments (pace Frege), and that these entailments need not necessarily be thought to be abandoned when those terms occur opaquely. After all, as Davidson vividly pointed out,³⁶ the natural reaction is to *resist* Frege's solution to the (putative) problem of the behavior of expressions in intensional contexts, imposing as it does what might well be thought of as a quite preposterous referential ambiguity as regards expressions in those contexts.

Keeping this Davidsonian attitude in mind, we can try to undermine some of the orthodox resistance concerning quantifying into intensional contexts by drawing an analogy between the so-called opaque contexts wrought by, for example, modal operators, and factive expressions, and propositional attitude contexts such as belief. The natural and intuitive expectation (and the one that Davidson makes much of) is that expressions that find themselves in the scope of such operators will not suddenly find themselves denuded of their usual or standard semantic value. On the face of it, at least, there is no reason why they should. A

demonstrative expression, for instance, has a certain function: it refers to a certain entity. And whether it succeeds or not is quite naturally supposed to depend on the way the world is, rather than on its location in a sentence.

Consider, to drive this home, a vacuous demonstrative. If embedding a demonstrative expression in an opaque context is sufficient to derail its usual reference to the extent that the Fregean picture maintains, what's to stop us supposing such an embedding is sufficient to make up for *failures* of reference? Why not suppose that a reference is generated for such expressions? Clearly this is unacceptable. The semantic role that is played by expressions outside of intensional contexts is most plausibly thought to persist when they are embedded. Inserting a vacuous demonstrative inside a belief context can't magically generate a reference for it.

Negation offers a straightforward commonsense case. Negating a sentence that contains a demonstrative expression doesn't inexplicably deprive the expression of its usual role of referring to a certain entity; the truth-value of the whole, familiarly, is a function of the truth-value of the parts. Existential entailments don't just vanish in these cases. Can't such commonsense apply equally to modal contexts as well? Is there any reason to suppose that the expression 'that *F* is *G*' won't do its usual semantic business - among other things, hang on to its customary

existential entailments - when it is embedded in the scope of the modal operator *It is necessary that* . . .? And similarly for sentences reporting factive attitudes. In the sentence '*C knows that that cat is an Angora,*' the embedded demonstrative expresses no fact without a corresponding cat: in the case of hallucination, no genuine state of knowledge can be ascribed. The key point to remember is that it is understandable to suppose that the principle of semantic constancy or uniformity ought to prevail - in fact, initially, it *does* prevail - with respect to expression outside and inside these contexts, and there is no in principle reason not to try to defend its holding for propositional attitude contexts - obstacles notwithstanding.

11. Weak Externalism: Formulation

Having commented on the presuppositions, we can now consider how it is that weak externalism can be formulated in terms of their combination. The next thing to examine is whether every sort of expression capable of occupying singular term position - descriptions as well as proper names and demonstratives - will admit of existential entailments. The essence of weak externalism, we might say, is located in this issue. For the orthodox understanding of existential import *denies* that general sentences (sentences of the form '*For all x. . . .*' and '*There is an x. . . .*') imply the existence of the referents of the expressions in

those sentences. And descriptions - according to Russell - are to be analyzed as general sentences. What will emerge is that although descriptive sentences *themselves* cannot be thought to have existential entailments, it is perfectly plausible that their *constituents* - the predicates that make them up - do.

Let us recapitulate. (4), above - a sentence containing a demonstrative singular term - can be true only if its logical implications are true; and one of these is an existential sentence. This is essentially equivalent to the claim that the belief in question is existence-dependent on the object of that belief: should the reference fail to obtain, no proposition is expressed: and if no proposition is expressed, so no proposition can be believed.³⁷

There is nothing particularly controversial about this. Obviously one wouldn't use a demonstrative in a belief-report unless one took it to refer: the report could only be true if the expression *did* refer. The idea is this: suppose Katie hallucinates a dog directly in her path baring its teeth and poised to spring. She no doubt *takes herself* to possess the belief that dog is baring its teeth and is poised to spring, which she would express by saying "That dog is baring its teeth and is poised to spring." Janet, observing her behavior, which includes looking frightened and clutching at her in a panic, turns to me, wondering what can explain Katie's behavior: Katie's *odd* behavior, given

that neither Janet nor I perceive any object in Katie's path. I cannot sincerely report Katie's mental state by uttering the sentence "Katie believes that dog is poised to spring," since *I* don't believe there is a dog there. Were I to attempt such an ascription, Janet would be fully entitled to ask - even testily - "WHAT dog?" In sum: I cannot report a belief with existential commitments when I possess disconfirming knowledge concerning the existence of the object of the belief. Segal, for one, makes this explicit with his own example (1989, 54-55):

Suppose that Orville, perhaps in the clutches of a mad scientist, is undergoing a visual experience that seems to present a particular olive. Orville, fond of olives and hungry, extends his hand in the appropriate direction. What is Orville doing? Orville, what are you doing? 'I am reaching for that olive.' When we try to describe Orville's action what we find ourselves saying is: Orville is reaching for the olive that he thinks is there . . . We say 'he is reaching for the olive that he thinks is there' because we are trying to say that he is reaching for an olive, which is what he thinks he is doing, without committing ourselves to there being an olive that he is reaching for.

This is the intuitive case for non-descriptive singular terms. The difficulties are thought to arise with respect to descriptive sentences.

Certainly the *meaningfulness* of 'There are *F*'s' doesn't depend on there being anything that is an *F* (whether or not such a sentence is true, however, does so depend). So

we have a clear distinction, on the Russellian picture, between expressions like proper names and demonstratives - which do depend on their reference for meaningfulness³⁸ - and quantified expressions, which do not. And this isn't so surprising, given that expressions like proper names and demonstratives are *singular terms*, whose very point is to refer to an object. Quantified expressions, as Russell notoriously pointed out, are *not* singular terms. The thing to realize at this juncture is that were there nothing to left to say about the existence-dependence of expressions, externalism in its most general formulation would not be established: for, at this stage, the only thing that emerges is the existence-dependence of non-descriptive singular terms on their objects. And there are many concepts that are not expressed by singular terms.

But there is something further to note. Consider descriptive expressions. Definite descriptions are not, according to Russell, proper singular terms, for they have the logical form of existential sentences. The sentence

(8) The author of *Waverly* was a man

is analyzed according to the Theory of Descriptions as

(9) There is one and only one *x* such that *x* wrote *Waverly* and *x* was a man

Now those familiar with the terrain here will recall that the canonical position concerning a descriptive thought is that one can possess, for instance, a descriptive belief without being *en rapport* (Kaplan 1968) with the object that satisfies the description: for instance, from

(10) Jerry believes that the most amusing cat in New York should go on the stage

there is supposed to be no clear inference to

(11) There is some cat in particular believed by Jerry to be the most amusing in New York and to thereby belong on the stage

since clearly Jerry can frame beliefs about the most amusing cat in New York and its theatrical potential by reasoning that cats come in different degrees of comic talent, and that there are cats in New York, so that there must be some cat or other who, by virtue of being the most amusing of the felines in that city, belongs on the stage. The canonical position regarding descriptive belief, that is, is that exportation is not a justified inference. But that would be to ignore the possibility that existential generalization might well apply to terms in predicate position as well to

those in singular term position. If this could be made plausible, then the case for the existence-dependence of even general thought on its objects would be at least *prima facie* secured.

Briefly, the idea is this. As we have said, for

(12) $(\exists x)Fx$

to be true, an individual (corresponding to the variable) must exist. But now consider the *meaningfulness* of (12). Certainly (12) can be meaningful in the absence of any individual. But if we consider (12) more closely, it's clear that there might be thought to be something further that needs to be true in order for (12) to be meaningful. And this is that the property referred by the predicate exist - after all, what makes (12) true or false is whether the world obliges with an object that *instantiates* that property.

What is of course presupposed in this line of thought is second-order quantification, argument for which is well outside the scope of my project here. Assuming, therefore, that second-order quantification (quantification over properties) is legitimate, then we may say that the meaningfulness, the *content*, of (12) presupposes that

(13) $(\exists P)(P = \lambda xFx)$

is true: what (13) says is that there is a property P such that P is identical to the property F . The content of (12) is thus existence-dependent on properties via the predicates it contains; and properties are extra-cranial entities.

This line of thought can be understood to extend a certain way of thinking about singular terms to predicates. Those familiar with that way of thinking will remember that Strawson (1956;1959;1974), and, later, Evans (1982), argued that if 'a' is a logically singular term, then for

(14) Fa

to express a proposition (to have content, to be meaningful), it must be the case that there is an x such that x is identical to a . Strawson writes (1974, 58):

. . . what about the very rare but not impossible case where there is an intended primary use of a name but . . . there just does not exist any real particular such that to indicate or specify that particular would be correctly to answer the question, 'Who/what is being named (or even mis-named) by the name?' . . . The answer I suggest is simple. If there is nothing which counts as command-of-the-name-as-then-used, then there is no proposition asserted, though the speaker by hypothesis thinks there is. . . .

Given, then, a putative utterance of a proposition of our basic class with,

say, a personal name in the 'a' place, we have it that a necessary condition of such a proposition being expressed by the utterance is that there should actually exist someone to whom the speaker is referring in uttering the name. . . .

The analogous move for predicates, as we have seen, requires that a second-order existential sentence be true; what (14) can be understood to express, that is, is that there is a property F such that a has F . And for this to be the case, it must follow, roughly speaking, that there exist a property F - although, of course, what won't follow is that anything *does* have F .

Let us summarize the basic points. Russell, as it happens, himself remarks (1988, 51):

. . . speaking generally, suppose we wish to say that the author of *Waverly* had the property ϕ , what we wish to say is equivalent to 'One and only one entity wrote *Waverly*, and that one had the property ϕ '.

And this serves to focus attention on a key idea. Descriptive sentences are made up predicates, as the analysis afforded by the Theory of Descriptions makes even more manifest. And while we can agree that general sentences do not bear a relation of existence-dependence to the entities over which the (first-order) quantifiers range, it does not follow from this that the predicate expressions that constitute those general sentences won't bear existence-dependence relations to extra-cranial entities.

Predicates are (assumed to be) open to second-order quantification, and second-order quantifiers range over properties.

After all, what is it that makes predicates meaningful? One perfectly straightforward answer (captured in presupposition (2), above) is *properties*; the properties that those predicates refer to. Predicate expressions can thus be thought to have existential entailments in the same way that singular terms do: such expressions depend for their meaningfulness on the existence of the properties to which they refer.³⁹ We have the basic picture in place for a *prima facie* case for the existential entailments of even general thought.

The picture that emerges from putting all of our presuppositions together is that, from (10), there is a perfectly plausible inference to

(15) There is an x (in particular) and there are properties F and G such that J believes that x instantiates F and G .

And, if so, it supports a notion of the existence-dependence of (both singular and general) beliefs on their objects. Given that, quite plausibly, (10) couldn't be true unless its logical implications were true - and some of its logical implications are the existential entailments of the expressions in predicate position, as well the existential

entailments of the expressions in singular term position - we are in a position to formulate weak externalism.

Mental states such as belief are relational - externalistically determined - in that they depend for their existence on their (mind-and-language independent) objects: the particulars and properties referred to by the expressions that make up the belief-sentence; the singular terms and the predicates. This existence-dependence is made manifest in the considerations concerning inferences from the notional to the relational forms of belief sentences. Given the admittedly controversial but perfectly acceptable intuition that what logical and semantic characteristics apply to expressions outside of intensional contexts ought uniformly to be supposed to apply to them when they are embedded, we have seen that there is no reason to suppose that both singular terms and definite descriptions - via the predicates that make up those descriptions - can't be thought to have existential entailments.

And if this is so, then descriptive content is no less existence-dependent on extra-cranial reality than is non-descriptive singular content - existence-dependent on extra-cranial *abstract* reality, to be sure, but extra-cranial nevertheless. McGinn sums up the point (1989, 41):

weak externalism . . . does seem to me pretty uncontroversial considered by itself . . . properties are precisely the kinds of item that *things in the world* instantiate. The item that has to exist for me to *think* (say) 'something

is square' is the very same item that has to exist in order for the material object in front of me to be square, viz., the property of being square. This property . . . has causal powers (when instantiated) that operate out there in the extramental world. It should strike us as more surprising (more significant) than it does that this very objective property also enters into the individuation of mental states: that what is objective and nonmental should enter into the very identity conditions of something subjective, a state of mind.

A few concluding - historical - remarks should suffice. One way to absorb the line of thought behind weak externalism is to recall the context in which it is located: that of a wider issue onto which the work on singular thought and descriptive content - prominent in the late 1970's and early 1980's - was directed.

A question can arise concerning the connection between externalism and the theory of descriptions: is it or is it not possible to avoid externalism with respect to particulars? According the theory of descriptions, all apparent singular terms can be analyzed away, so there is no reason to suppose that some thoughts necessarily involve reference to a (head-extraneous) particular: all thought is general thought.

Now, apart from the objection that the theory of descriptions may not apply to all denoting singular expressions, there is the following. What motivates the view that externalism should be denied for particulars? It seems

that a determined *internalist* about the mind might well try to deny the existence of singular thought - since the necessary condition on the existence of a singular thought is that the worldly object of that thought exist - and think to succeed in this suppression by characterizing all thought as *general*. But this would be to fail to notice that descriptions themselves might arguably be indirectly linked to extra-cranial reality. Descriptions are composed of predicates: expressions that, plausibly, refer to properties. The internalist (masquerading as a description theorist) determined to deny an extra-cranial link between thought and its objects would then have to presuppose a particular ontological view about the status of properties - that these are mentalistic or in some way internal - in a rather desperate move to avoid externalism about even general thought.

Ultimately, the issue of whether or not there exists singular thought is independent of the debate between externalism and internalism, since singularity need not in itself imply extra-cranial content links (suppose you thought the reference of all singular expressions were sense-data). However, there is a convergence, since it will be *sufficient* for externalism that there be singular thoughts about non-mental entities. The point is, however, that even if it could be argued that there were no singular thought about non-mental entities, externalism would not

necessarily thereby be avoided: externalism could still be claimed with respect to descriptive content, by raising issues concerning the ontological status of properties. And this brings the sketch up to date (for more, see Blackburn (1984), Evans (1982), McCulloch (1989), and McGinn (1989), among others).

12. Weak Externalism and Propositional Constituency

Weak externalism's claim that mental states are necessarily individuated by mind-and-language independent reality - that their existence-and-identity-conditions are those of objects and properties that are the reference of those states - can be formulated in a slightly different way, according to a line of argument associated with Russell (1918), and developed by Evans (1982) and Kaplan (1972, 1978), among others. This is the view, roughly speaking, that states of affairs have constituents, and, given that propositions represent states of affairs, it is quite plausible that propositions have the very same constituents that do the states of affairs they represent. It follows from this that the existence conditions of a belief content are those of the state of affairs the content represents: for the having of a belief is to bear a certain complex psychological relation to a state of affairs.⁴⁰ Let us review the basic architecture of this view.

Kaplan notes (1979, 212):

There is also an important - though less often noted - difference between Frege and Russell regarding the structure of intensional entities. According to Russell, an individual may be an immediate element of a proposition. In fact, certain *atomic propositions* consist of just individuals and attributes (or relations); whereas, for Frege, the immediate elements of a proposition must themselves be intensional entities of one sort or another.

This is borne out in the correspondence between Frege and Russell on the subject of propositional constituency (Frege, 1980, 163;169):

Truth is not a component part of a thought, just as Mont Blanc with its snowfields is not itself a component part of the thought that Mont Blanc is more than 4000 metres high. (Frege to Russell, 13.11.04)

Concerning sense and meaning, I see nothing but difficulties which I cannot overcome. . . I believe that in spite of all its snowfields Mont Blanc itself is a component part of what is actually asserted in the proposition 'Mont Blanc is more than 4000 metres high.' We do not assert the thought, for this is a private psychological matter: we assert the object of the thought, and this is, to my mind, a certain complex (an objective proposition, one might say) in which Mont Blanc itself is a component part . . . This is why for me the *meaning* of a proposition is not the true, but a certain complex which (in the given case) is true. (Russell to Frege, 12.12.04)

Kaplan (1979, 218) adapts the Russellian conception of a proposition and makes a distinction between *singular*

propositions (those propositions which contain individuals as immediate constituents; 'This is blue') and *general propositions*, which do not contain individuals as constituents ('All men are mortal'). Kaplan's view - sometimes known as the Theory of Direct Reference - is that a propositional attitude like a belief is a relation to a proposition, whose constituents are the reference of the expressions of the sentence that expresses the belief: for singular propositions, objects, for general propositions, properties. It is true that Kaplan himself does not extend his points about propositional constituency to general propositions,⁴¹ nor does he draw the implication for possession of thought or the lack of it, should the relevant objects (properties) fail to obtain, but the implication is obvious. If to be in belief-state *b* one is held to be in a relation to a proposition whose constituents are object *o* and property *P*, and *o* and *P* fail to exist, then one is not related to a proposition: so one fails to be in belief-state *b*.

Evans's work on the subject of what he calls Russellian thought (1982) parallels Kaplan's but makes the existence-dependence of singular propositions on their objects completely explicit. For those unfamiliar with Evans on the subject of Russellian thought, the following is a precis of the position: content involving non-descriptive singular terms cannot be possessed in the absence of the

object.⁴² The possession of descriptive thought is subject to conditions that are a little less uncompromising, as descriptions are analyzed according to Russell's Theory of Descriptions.⁴³

According to Evans (1982, 44-6; 30-1):

Russell . . . held a theory of thought which incorporated the principle that it is not possible for a person to have a thought about something unless he knows which particular individual in the world he is thinking about . . . Given Russell's restriction, a situation can never arise in which a subject thinks that he is having or expressing a thought about an object while failing to do so; and this was a possibility which Russell very much wished to rule out, because it seemed to him incoherent. . . However, there does not seem to me to be anything incoherent in the idea that it may be, for a subject, exactly as though he were thinking about a physical object (say) which he can see, and yet that, precisely because there is no physical object he is seeing, he may fail to have a thought of the kind he supposes himself to have. It is not part of this proposal that his mind is wholly vacant . . . The claim is simply that there is a kind of thought we sometimes have, typically expressed in the form 'This G is F', and we may aim to have a thought of this kind when, in virtue of the absence of any appropriate object, there is no such thought to be had.

Anyone who is attracted by a Russellian view of a class of singular terms must always attempt this further task: the task of explaining why, when a member of the class is empty, there is such a strong impression of understanding, communicating, and thinking.

Now, weak externalism extends, to general concepts, the idea of the identity-and-existence-dependence of content

on the world outside the subject's skin.⁴⁴ There is no *prima facie* reason to rule out properties being themselves constituents of propositions, (assuming there are any such things as properties). Russell (1912, 58, emphasis his) notes:

Many universals, like many particulars, are only known to us by description. But here, as in the case of particulars, knowledge concerning what is known by description is ultimately reducible to knowledge concerning what is known by acquaintance . . . every proposition which we can understand must be composed wholly of constituents with which we are acquainted.

Of course, if one suffers from a deep-seated scepticism about the existence of properties, one will fail to be persuaded by the unzipped realism presupposed in weak externalism's Russellian conception of a general proposition. Now, it is no part of my project to argue for realism about properties; a remark by Field (1981, 110), however, sets the tone⁴⁵:

Some people may feel that there is no ontological gain in quantifying over properties rather than over propositions. Such a person should read Putnam (1975f). Putnam makes a good case (a) that quantification over properties is needed in science, and (b) that properties are quite distinct from meanings, in that two predicates like 'x has temperature 210°C' and 'x has mean molecular energy 10⁻²⁰ joules' can turn out to stand for the same property even though they clearly differ in meaning.

Here, however, we can make a few points about abstract

objects and weak externalism, in anticipation of the charge that externalism is not relevantly supported by the claim that extra-cranial abstract reality has a content-individuating role.

13. Externalism and Abstracta

There is, of course, no reason to lump all abstract objects together from the point of view of their potential role as content-individuating items. We must distinguish, say, between Fregean senses and properties. Fregean senses are, of course, not in the head; Frege was himself adamant that they be considered objects in the public domain, so as to serve as media through which cognitive information could be transmitted. Now, it is an interesting question in what respect, if any, content individuated by way of Fregean sense can be considered externalistic in the sense we are interested in here (we might perhaps call that notion Very Weak Externalism, if we could bring it off),⁴⁶ but it is one that would take me too far afield here. I will confine myself to a few suggestive remarks.

It seems to me to be a gross distortion of Frege's canonical position to claim that Fregean senses belong to the *world* in a sense of world illuminative of the notion of externalism we are interested in formulating here. It may be useful to remember the characterization of Fregean sense given by Barwise and Perry (1982, 4), who attribute to Frege

the notion of a "third realm," a realm "neither of ideas nor of worldly events;" the realm to which sense can be thought to belong. Barwise and Perry may get this from Popper (1972), who puts the point in terms of "worlds": he distinguishes the world of physical objects or physical states; the world of states of consciousness; and, finally, the world of objective contents of thought. We can adapt his usage and speak of the 'world of sense' as distinguished from the 'world of reference.' Both of these 'worlds' can be thought of as mind-and-language independent,⁴⁷ but only one of them, according to this line of argument, will have a genuinely externalistic role to play in the individuation of mental content.

To see this, one must keep in mind the notion of externalism that is at stake here. Externalism, as we have been formulating it, is a claim that the nature of mind is essentially determined by something usually thought of as utterly different in character - states of affairs; the movements and permutations in objects and their properties. What makes this an interesting and innovative claim is precisely that the mind, on this view, is conceived of as a set of relations to the world of objects and their properties: the ordinary world, the one that engages the interest of scientists and ordinary people.⁴⁸ This flies in the face of the usual conception of mind: as a substance, as autonomous with respect to the world of things it is

presumably directed onto in its states of perception of knowledge, as directly, introspectibly, infallibly accessible.

Now, to say that the essential character of mind is determined by a relation to Fregean senses simply doesn't have the same kind of bite, and there is a plausible way to explain that: following our usage above, Fregean sense does not belong to the world of reference (by definition). Fregean sense is a semantic device, introduced to explain certain *psychological* phenomena: for instance, why it is that in spite of the fact that the reference of two expressions *a* and *b* is identical, different cognitive information is expressed by the sentence '*a* is *a*' than is expressed by the sentence '*a* is *b*.' Fregean sense is supposed to explain why knowing the truth of '*a* is *a*' is trivial, while knowing the truth of '*a* is *b*' is informative. Now, none of this has anything to do with the world of interest to scientists and ordinary people: the world, as it is sometimes characterized, as it is in itself. States of affairs are not determined by the, say, causal interaction of their Fregean senses (it is difficult to know what this would even mean) - they are determined, if by anything, by the permutations undergone by objects and their properties. Objects, likewise, do not instantiate Fregean sense: they instantiate properties. There is a plausible distinction to be drawn between the role that properties could conceivably

play in the externalistic individuation of mental content and the role that senses might play: to deny this is to be perverse.

It cannot be emphasized enough that clinging to a TE formulation of externalism will obscure one's grasp of these points. For example, an objection might be raised that properties, being abstract objects, are necessary existents, so there is no such case as one where two individuals inhabit a different environment: their *abstract* environment will always be the same; thus, no externalism. Clearly, this is to confuse the formulations. Of course, the notions of 'environment' or 'world-involving' or 'world-dependent' are not always understood in the same way in the literature. Some take them to mean the causal context - the familiar twin earth story - but others take them merely to mean what Pettit (1986) calls the "surround" outside the subject's skin. This demarcates the 'environment' or 'world' in a broader way than does the twin earth story, and supports the view that 'environmental' or 'worldly' entities need not necessarily exclude, for example, abstract entities such as properties. This is a rather complicated issue, which I believe is illuminated by distinguishing properties from other abstract objects in terms of their causal powers and their role in explanations about the physical world (See, among others, Armstrong, 1983, 1989; Dretske, 1977; Haugeland, 1980; Shoemaker, 1984; Tooley, 1987). I consider

this in more detail in chapter 3.

We might further remark that there are conceptions of universals which link them to their instantiation (cf. Aristotle: "For the species is synonymous with its individuals"; *Topics*, 154a18); surely not every world will instantiate all properties. The thing to keep in mind is that that externalism is a theory of mind from which a determined Cartesian about the mind would recoil; what makes it so repulsive to the Cartesian is its adherence to a genuinely constitutive necessary relation between mind-independent reality and the the mind; a relation analogous to that, say, between a set and its members. It is a necessary condition - but not sufficient - on candidates for individuation that they be mind-independent; and properties, instantiated or not, satisfy that condition.

14. Conclusion

We have been concerned, in this chapter, with formulating a notion of externalism that captures the idea that there is inextricable relation between independently obtaining states of affairs and the mind. We have argued that there is natural correlation between issues to do with meaning and reference and those to do with the fixing of mental content by head-independent reality. We have shown that the Twin Earth formulation of externalism - that content fails to supervene on inner states - though

prominent, fails to reckon with the idea that there may be content for which failure of supervenience is not an option; as the price of popularism is, familiarly, precision, this may perhaps come as little surprise. The exposure of the Twin Earth formulation as very likely limited to a few kinds of content raises the possibility that there is a more general formulation of externalism that more accurately captures the general intuition that Twin Earth cases were meant to support.

We have distinguished externalism into two strands: strong and weak; and have argued that weak externalism can be established by the use of familiar semantic principles concerning, among other things, expressions and their existential entailments. None of this is without controversy, of course. But a case has been put forward for a general and necessary connection between content and the extra-cranial environment, a view whose interest stems from its direct challenge to the traditional Cartesian model of the mind as an entity of an incomparable and immaterial sort, permutating in splendid isolation from the world of objects and properties.

The consequences of externalism for psychological explanation is a familiar issue. In Chapter 2, I argue that the repercussions for explanation are more widespread than has been hitherto supposed, given our more general formulation of the notion of externalism.

NOTES TO CHAPTER 1

1. The notion of mental states as existence-dependent on the extra-mental environment can be found originally in Russell (1988, 175-281;285-322).

2. One that, as I intend to show, is more or less spurious. In fact, however, there seems to be a split on the understanding of 'externalism' depending on which side of the Atlantic one is on. In England, the issues center around singular or object-dependent thought; in the US, Twin Earth cases - failures of supervenience - are the dominant criteria for establishing externalism (those on the west coast (Kaplan, say), are closer to the British, however, in their formulations). A glance at the references to this work will give some idea of the division.

3. The condition is a necessary one, since it is otherwise difficult to see how one could establish a decisive distinction between a theory of mind that claims that the world, as it happens, *corresponds* to the mental states directed onto it; and one that claims that the fact that mental states have intentionality is partly because the world is a *constituent* of mental states.

4. More will be said about this controversial claim below: see the discussion of Evans (section 8).

5. I subject this to detailed scrutiny below in section (6).

6. Putnam is sometimes credited with being the first to raise the issue of externalism, which has for one reason or another come to mean, in the minds of some, the view that content fails to supervene, but the fact is that his concerns about whether meaning was or was not in the head is a clear historical progression - with a slightly different emphasis - from the concerns of Russell about the nature of the proposition. It was Russell, actually, who first argued against Frege that the constituents of propositions were the

very objects of the proposition, as against Frege's view that the constituents of propositions were, rather, a mediate, semantic entity. Naturally the Russellian view has interesting consequences for the individuation of content and for belief-ascription: what if the objects don't exist? What if two people have exactly the same internal makeup but are causally related to two completely different objects? What if the objects are unfailingly mind-and-language independent, whether they are contingent or abstract? It is this extremely general issue that I take as the context against which the critical remarks in this chapter need to be understood.

7. Psychological states that are methodologically solipsist can be understood to be psychological states in the narrow sense; Putnam makes the point and Fodor generates an industry: see Fodor (1980, 63-109).

8. See McGinn (1989, 31).

9. The title of this paper - a compressed version of "The Meaning of 'Meaning'" - says it all. But, just for the record, note Burge's remarks (1979, 117, n. 2) in summarizing his position and its connection to that of Putnam's:

Putnam himself does not give quite this argument . . . This is partly just a result of his concentration on meaning instead of propositional attitudes.

10. The fact is that even for content for which TE cases are usually thought to apply, a careful examination of the facts reveals that the canonical case - varying the environment while keeping internal states the same - doesn't always make a perfect fit. Take for instance beliefs about oneself. Using 'I,' two speakers utter the sentence "I'm hot." Their internal states are the same, but it is stretching a point to say their environment has varied - what has varied is the identity of the speakers. This, I suppose, is sufficient for the purpose at hand; but it isn't a canonical TE case.

A separate but related point is this. Once you earmark for the reference of an expression (or a mental state) a necessary role in the determination of content, then the way is clear to claim that content is existence-dependent on its objects. So, in the case where two people believe a horse is snorting behind them, and one of them is hallucinating, then

only one of them is in possession of the belief that that horse is getting nearer - the relevant (individuating) object of the belief being a figment. This kind of case can be understood in terms of non-supervenience: having the belief in question doesn't supervene on inner states - but there is no necessary connection; putting it this way gets the point across but involves none of the canonical TE apparatus.

The right thing to say about this (given that the examples involve indexicals, which Putnam argues are the other kind of content for which TE cases can be erected) is that even for the kind of content for which TE cases can be erected one should realize that TE itself is unnecessary, as is varying the environment in some dramatic sense. These points ought really to eradicate any lingering idea that TE is somehow necessary for making claims consistent with externalism. The really interesting question concerns content for which TE cases CAN'T be erected: can content both supervene and be thought to be externalistically determined? We shall see below that the answer is yes.

11. Putnam (1975a) himself appears to make the essential points in the section entitled "A socio-linguistic hypothesis" (227-229).

On a different note, it is worth pointing out, for the sake of accuracy, that Putnam does not, in (1975a), actually draw the conclusion that the *mental states* of Twins differ when their environments do: his view in 1975a, summarized by McGinn (1977, 531), is

that two speakers, or two linguistic communities, could use (phonetically identical) terms whose extensions in their respective languages were disjoint and yet be in the same mental states with respect to the terms and their extensions.

Burge (1979, p.117, note 2) concurs:

[Putnam] remarks in effect that the subject's *thoughts* remain constant between the actual and counterfactual cases (p.224). In his own argument he explicates the difference between actual and counterfactual cases in terms of a difference in the extension of terms, not a difference in those aspects of their meaning that play a role in the cognitive life of the subject. In my view, the examples do illustrate the

fact that all attitudes involving natural kind notions, including *de dicto* attitudes, presuppose *de re* attitudes.

12. This notion was made much of later by Kripke, (1980), Lecture III.

13. It seems that Burge himself seems to think such a condition is imposed; see Antony (1993, 247-48) who accepts Burge's own valuation of the matter, but goes on to challenge it:

Of all Burge's thought experiments those alone [in 'Individualism and the Mental'] support his thesis that social relations are essential to the nature's of one's thoughts . . . Burge does provide an account from which it follows that Yolanda's actual and counterfactual thoughts differ - his story that 'language-community membership' is essential to the natures of one's thoughts.

14. Certainly a number of commentators proceed as if this is Putnam's claim: see almost anyone British who writes in this area. Anyway, as we have said above, the interest of externalism as a general characterization of mental content would be significantly diminished if the claims in question weren't thought to be imposing necessary conditions on concept-possession.

15. Here the use of the term should be understood in Burge's sense. Note that 'individualism' is used in at least two ways in the literature. Burge (1979, 103) originally uses it to

apply . . . to philosophical treatments that seek to see a person's intentional mental phenomena ultimately and purely in terms of what happens to the person, what occurs within him, and how he responds to his physical environment, without any essential reference to the social context in which he or the interpreter of his mental phenomena are situated.

Fodor (1988, 30), on the other hand, uses 'individualism' to apply to: "standards of individuation according to which my Twin and I are in the same mental state. . . . "

This might be a good place to make a few terminological points. It has become standard to use 'anti-individualism' to mean 'externalism' at least in some sense. This seems to me to add a certain amount of confusion to an already complicated set of issues. We will take 'individualism' and 'anti-individualism' to refer to issues having to do with what McGinn (1989, 2, n.5) calls the subject-other relation; and 'internalism' and 'externalism' to refer to issues having to do with what he calls the subject-object relation. As McGinn puts it:

The question whether minds can fix the content of other minds is really a very different question from the question whether the extramental world can fix what holds of minds: different 'interfaces' are being considered in the two cases.

16. Pettit and McDowell (1986, 1-15) use this notion in their formulation of Putnam's thesis.

17. Realism about moral and aesthetic concepts is here presupposed.

18. Burge (1979), despite a number of remarks about color concepts, doesn't appear to face up to this point. He considers the case of color ranges (p. 82):

People sometimes make mistakes about color ranges. They may correctly apply a color term to a certain color, but also mistakenly apply it to shades of a neighboring color. When asked to explain the color term, they cite the standard cases . . . but they apply the term somewhat beyond its conventionally established range. . . .

That's fine as far as it goes. But then he goes on:

The error is linguistic or conceptual . . . It is not an ordinary empirical error. But one may reasonably doubt that the subjects misunderstand the

dictionary meaning of the color term
. . . we can imagine that 'red' were
applied as they mistakenly apply it. In
such cases, we would no longer ascribe
content-clauses involving the term 'red'
in oblique position.

This seems to suggest that when a subject's mistake
about color-ranges goes so far as to lead him to think a
color term applies to two colors (a sort of extreme color-
range mistake, perhaps), then the conditions for the
thought-experiment break down.

19. He who remains sceptical is invited to try to
erect a Burge-type case for concepts of belief and desire,
or concepts of pain and pleasure. Surely we want to deny
that it is possible to be wrong about whether one has a
belief - as opposed to a desire. Yet, if on TE the words
'belief-state' are used by the community to apply to
desires, then my twin and I are in the same state - that of
desire - which I *inaccurately* ascribe to myself as 'belief-
state,' since on Earth, the community uses 'belief-state' to
apply to belief states. This is really quite implausible.

20. McGinn further points out that even this
characterization of the conditions on formulation assumes
that 'personhood' is not environmentally determined -
something that might be open to question.

21. I have no particular views here: I am merely
spelling out the logical possibilities for the sake of
accuracy. The fact is, you could take the view that brain
states are individuated by their function or purpose: if
you're a teleologist about brain states then you're an
externalist about such states, and the preconditions for TE
cases might not be met. See Millikan (1984) for the basic
picture; see also Fodor's (critical) discussion (1988, 31).
I discuss teleological views of content below: section x,
(e).

22. I don't want to anticipate my later remarks too
much at this ground-clearing stage: suffice it to say that
McGinn's distinction between strong and weak externalism -
the distinction in formulations of externalism introduced at
the beginning of the chapter - applies to perceptual
content. Qualitative states are weakly, but not strongly,
external, as he argues (1989, 58-100). Externalism about
perceptual content is a thorny issue, and not one that is

relevant to my concerns in this thesis.

23. It must be said that the original formulations of externalism *qua* externalism (and many subsequent) do not avail themselves of the notion of supervenience - Putnam (1975a) most certainly never mentions it, and neither do Evans, McDowell, or anyone else working in *that* tradition of thought about externalism. This concept has crept into the externalist literature and become entrenched - at least in some parts of the world - mainly, I would say, as a result of Fodor's arguments (1980, 1988) concerning content and explanation. But now there is every reason to suppose that the notion of supervenience might be actually somewhat ambiguous.

24. Say, for example, the relation of the moral on the descriptive, the aesthetic on the physical, and the modal on the actual.

25. But let's be careful here. Those who are unfamiliar with the literature concerning weak and strong externalism might easily simply take 'relational' to be equivalent to 'bears causal/environmental relations to . . .' This is of course because such persons confuse TE with externalism. We stipulate rather that "relational" be understood as generally as possible: to *include* causal/environmental relations, but not to be exclusively restricted to those.

26. Much more about this below.

27. These arguments are by no means original to me: McGinn, adding to certain key ideas from Russell, Strawson, and Evans, makes them in his 1989, chapter 1. More on this below.

28. It must be said that there are issues that could arise here that are extremely complex and raise all sorts of difficulties; for example, the remarks below will touch on exportation, the reference of predicates, Russell's theory of descriptions, opacity, vacuous singular terms, fictional discourse, negative existentials and *de re* modality. I resist getting drawn into the controversies that surround these fearsome issues. All I am concerned to do here is to examine the ways in which some of these notions and that of externalism in its most general formulation might be seen to

converge.

29. We have already pointed out that nearly all of those who write on the subject of externalism put it in terms of object-dependent thought - anyone working on the subject, that is, in England or the on West Coast. It really is only a handful of people who put this issue in terms - exclusively - of failures of supervenience. It remains to be noted that most of the work done with respect to object-dependent thought concentrated on (non-descriptive) singular terms; for a long time, for instance, demonstratives were all the vogue (see Evans, 1982; McDowell, 1977; 1984 for starters). McGinn 1989 was the first to extend the notion of object-dependent thought to the *predicates* involved in attitude ascription.

30. The issue of vacuous singular terms is one with familiar complications, comment on which is mercifully not part of my project here. A comment on the semantic behavior of indexicals like "I" and "now" might be useful, however: from the fact that it is difficult to erect a case with a vacuous referential use of "I" it won't follow that the *semantics* of this kind of demonstrative expression differs from the others: it simply represents a sort of limit case.

31. McGinn (1989, 37):

The term 'that dog' . . . has need of a demonstrated dog if the sentence is to succeed in stating a fact.

32. The sceptic about denotation and existence-dependence is invited to examine the history of these notions. The place to start is Russell (1988, 39-56; 175-282). If that isn't sufficient, Evans (1982) is worthwhile. Finally, Kaplan (1972) is useful.

33. In particular, I will not consider issues to do with vacuous names, fictional discourse, negative existentials, or *de re* modality. These are obviously interesting but too complicated to be treated in this paper, and, in particular, not relevant to my project, which is to formulate externalism as generally as possible. I will, however, have something to say about descriptive singular terms. For a survey and treatment of the problems associated with these issues, see Evans (1982).

I would make one point however, concerning these

difficult cases. As those familiar with the literature will know, it is sometimes argued that these cases show that there cannot be any such thing as object-dependent or singular thought. One line of argument against this one might proceed as follows: there has got to be a fallacy involved in the argument that claims that because

(0) It is not the case that Hermes is mortal

can be true without there being a real god that is mortal it must follow that

(0') Alan is disgusted

does not presuppose the existence of a concrete individual. If this argument were correct, that would mean that no sentence ever presupposed the existence of a concrete, real individual. And it's doubtful that the reference relation is in that much trouble.

34. An analogy can be drawn, as is well-known, with modal contexts.

35. See at the very least, Kaplan (1968).

36. For more on this, see Davidson (1984, 93-108); also chapter 3 below.

37. Once more, the sceptic is invited to review his or her Russell, Evans and Kaplan. I emphasize this since resistance on this point will completely interfere with grasp of the essence of this chapter.

38. Of course, this is a slight misrepresentation of the conservative Russellian view of what counts as a proper singular term. As those familiar with Russell will recall, the only expressions that count as *logically proper names* are the expressions 'this' and 'that' - and these only when they are used to refer to one's own sense-data.

39. A very good source for views on the subject (as well as critical commentary and a defense of property realism) is Armstrong (1978, Vol. I); see also Bealer (1992). It so happens that Burge makes a relevant remark (1982, 120):

As I mentioned earlier, one might hold that 'water' names an abstract property or kind and that attitude attributions typically attribute *de re* attitudes of the kind. . . .

40. Classic sceptical arguments, say about the existence of the external world, proceed by noting that one can indeed instantiate the belief that the table is in the room even if there isn't a table (or a room). This is because there would appear to be a distinction between having the belief that *p* and having the *veridical* belief that *p*. The literature opposed to this view does indeed claim that for some content (non-descriptive singular content), an ascription cannot be made unless it is *veridical* (see Evans, 1982, among others, for an extended discussion). Naturally this is controversial.

41. But Salmon does. See his (1982), particularly chapter two.

42. The issue concerning the possession of negative existential thought is a familiar complication; Evans has much to say on the subject, and the reader is referred to his (1982). My desire is simply to make use of some basic concepts from Evans, not to treat his views to extensive critical scrutiny.

43. And so (1982, 47; 43):

. . . Russell held, and I think he was entirely right to hold, that where a clear descriptive condition exists for something's being the referent or denotation of a term, a quite determinate truth condition is associated with sentences containing the term, whether or not it is empty; the sentence is true just in case there exist something which uniquely satisfies the condition, and which satisfies the sentence's predicate. . . .

Where ' ϕ ' is a coherent description, perfectly clear conditions for the truth of the sentence 'the ϕ is *F*' (and

thereby equally clear conditions for its non-truth) have been laid down: the absence of a satisfier of the description is no obstacle to someone's correctly understanding an utterance of the sentence as having these truth-conditions. A thought may be conveyed; and a belief (that the conditions are satisfied) may be induced. I do not believe that it is open to us to hold, with Strawson, that someone who utters a sentence containing an empty description has said nothing (expressed no proposition). . . .

44. It cannot be emphasized enough that basic ideas must be kept clear in order to avoid getting sidetracked on to orthogonal issues. The claim that reference is a necessary condition on meaning/content is not equivalent to the claim that it is a necessary and sufficient condition. (Worth remembering here, as an example, is that the Evans/McDowell view of Russellian thought does not preclude sense (See Evans, 1982, 39;51)). As a purely logical point, it is surely obvious that from the fact that reference is taken to be a necessary propositional constituent, it will not follow that reference is taken to be the only propositional constituent. The claim that propositions are held to contain worldly items is not in itself a claim that worldly items is all they contain. In a word, it won't do to mix up the claims of a Direct Reference Theory with, for instance, the more conciliatory claims of dual componency theories of content.

Of course, the question: "In virtue of what are the propositions expressed by

(1) Batman is Batman

and

(2) Bruce Wayne is Batman

distinguished, given that a) Catwoman believes (1) and doesn't believe (2); and b) on the account under consideration, these propositions have the very same referential constituents?" is a perfectly good question. It is not one whose answer is essential to my project, and I do not anticipate delving into the vast literature on the subject, as it is to one side of my interests.

It is worth pointing out, however, by way of suggestive remark, that:

(1) propositions can be claimed to have the same referential constituents without it following that they must contain the very same intensional constituents.

(2) We can draw a plausible analogy with perceptual states such as seeing an object, to support the claim that propositional states can share the same necessary conditions without sharing the same sufficient ones.

(3) It won't do to understand 'containment' or 'constituency' too crudely: it is best to understand these as a way of talking about the necessary conditions on content.

45. The account of analyticity given semantic externalism formulated in chapter 3 will rely on a firm distinction between concepts and properties. Some remarks to challenge mentalism about properties will be made there.

46. On further thought, however, it seems likely that we shall have to deny that any such notion of externalism is defensible. One might be tempted to think that there is variety of externalism formulable in terms of the content-individuating role of any and all abstract entities, these not being in the head. What appears to be at stake here is the claim that it is sufficient to generate externalism that the phenomenon in question involves a relation to an abstract entity: the having of belief *b*, for instance, in that it involves a relation to some abstract entity, thus implies externalism about belief-individuation.

But this claim appears open to a glaring reductio. Take, for instance, an object's being square. An object's being square is plausibly a function of a relation between the object and an abstract object - the universal 'square.' But surely no one in their right minds would want to argue that since an abstract object is involved in something's being square then we have to be externalists about object-individuation. So, analogously, for claims about mental state individuation, for this sort of view implies that you are an externalist in virtue of instantiating any property: like the property of being in pain, or the property of instantiating neurophysiological property *p*, both commonly considered quite internal properties. What's worse, this sort of view implies that you are an externalist in virtue of believing in beliefs at all: *being the belief that snow is white*, is, after all, an abstract entity.

47. Although, as Fregean exegesis will attest, there is some controversy over whether Fregean senses are indeed language-independent.

48. McGinn (1989, 18, n.24) makes the point that there is an interesting analogy between the externalistic determination of mind and the relational determination of space.

CHAPTER 2

EXTERNALISM AND ITS CONSEQUENCES FOR PSYCHOLOGICAL EXPLANATION

. . . Koznyshev was not alone. With him was a famous professor of philosophy who had come from Kharkov for the express purpose of clearing up some difference that had arisen between them on a very important philosophical question . . . They were discussing the fashionable question whether there was a dividing line between the psychological and physiological phenomena in human activity; and, if so, where?

Tolstoy, Anna Karenin

1. Introduction: Causation and Explanation

A number of related questions arise concerning the occurrence of an event. Two of the most predictable, and familiar, are "what caused this event?" and "what explains the occurrence of this event?" That these questions can be distinguished is something that has been very thoroughly examined by Davidson (1980), among others, whose remarks on the subject serve as a useful background. We will briefly rehearse a few points about causal relations and their descriptions, then move on to consider the consequences of semantic externalism for mental events and their explanation.

Davidson, familiarly, does not directly consider the nature of causal relations¹, but focuses instead on the logical form of causal sentences. The sentence:

(1) The explosion caused the evacuation of the World Trade Center

expresses a two-place relation, with the singular terms 'the explosion' and 'the evacuation' referring to particular events. Now, in that this sentence states a causal relation between two events, it is referentially transparent. That is, replacing the two singular terms with other co-referential singular terms, as in

(2) The act of terrorism linked to the blind Islamic cleric caused the biggest catastrophe in New York in 1993

results in no variation in truth-value. So, briefly, it appears correct to say that causal relations hold between events and do so no matter how the events are described.

The same thing, however, cannot be said for explanations. When the question concerns the *explanation* for the evacuation of the building, rather than the relation between the evacuation and its causal antecedents, then it appears to be the case that only *some* purported explanation sentences will be adequate to the job, not all.

A natural reply, when asked why the building was evacuated is, "an explosion went off." The property of being an explosion here almost certainly can be said to have an explanatory role. Explosions involve disintegration of

materials and the like, which appear to reveal the mechanism at work behind a sudden mass exodus out of a building. Consider, on the other hand, whether the following reply would be satisfactory: "the building was evacuated on account of the occurrence of an event linked to the blind Islamic cleric in New Jersey." This reply is unsatisfactory in that it seems quite beside the point, given that we want to explain the connection between the explosion event and the evacuation event. That these events had something to do with the cleric does not seem to get to the root of the question: after all, lots of things are linked to the cleric every day, not many of which result in evacuations of buildings. His merely being linked to an event, that is, isn't sufficient to bring about the evacuation of a building.

The upshot is that not every description of an event - even if it is a true description - will refer to the properties of the event that have a role in an explanation involving that event. Metaphysically speaking, this comes down to the fairly uncontroversial claim that for any event, many things (even infinitely many things) are true of it, but not all of those things will explain it or its effects. In sum, not every property of an event has an explanatory role.

Clearly, however, if one is interested in explaining an event, then one needs some criterion to narrow down those

properties of an event that do have an explanatory role. A criterion that is widely thought to be useful involves laws and counterfactuals.²

2. The role of laws in scientific explanation

Hempel is noted for, among other things, a classic treatment of the role of laws in explanation. His *deductive-nomological* model (D-N model) of scientific explanation (1965, 336-337):

. . . effects a deductive subsumption of the explanandum under principles that have the character of general laws. Thus a D-N explanation answers the question 'Why did the explanandum-phenomenon occur?' by showing that the phenomenon resulted from certain particular circumstances, specified in $C_1, C_2, \dots, C_k$ [sentences describing the particular facts invoked], in accordance with the laws $L_1, L_2, \dots, L_r$ [the general laws on which the explanation rests]. By pointing this out, the argument shows that, given the particular circumstances and the laws in question, the occurrence of the phenomenon was to be expected; and it is in this sense that the explanation enables us to *understand why* the phenomenon occurred. In a D-N explanation, then, the explanandum is a logical consequence of the explanans. Furthermore, reliance on general laws is essential to a D-N explanation; it is in virtue of such laws that the particular facts cited in the explanans possess explanatory relevance to the explanandum phenomenon.

Hempel's criterion illustrates one paradigm of the nature of explanation: that explanation in science is concerned to uncover the mechanisms that operate with

respect to the objects and events that we observe. These mechanisms are understood as the links between causes and effects of phenomena, and science is understood to offer causal explanation (Hempel 1965, 347-349):

An explanation of a particular occurrence is often conceived as pointing out what 'caused' it . . . But causal attributions . . . presuppose appropriate laws . . . And by virtue of thus presupposing general laws which connect 'cause' and 'effect,' causal explanation conforms to the D-N model . . . And the law tacitly implied by the assertion that *b*, as an event of kind *B* was caused by *a* as an event of kind *A* is a general statement of causal connection

In sum: explanations are deductive arguments, the conclusion of an explanation is a sentence that describes the phenomenon to be explained, and the premises of the argument contain a natural law.

Hempel's criterion, although it relies heavily on the notion of a law, does not in itself offer any characterization or explication of the notion of law, as he himself notes (1965, 340-345). This is an issue best left to one side here; one or two remarks, however, can help to shed light on the issue of the explanatory role of properties of objects and events.

Properties customarily thought to characterize a law are, first, its generality, and second, its being counterfactual supporting. Take the statement "All the cats in my apartment are black and white." Does this entail the

conditional "If x were a cat, and x were in my apartment, then x would be black and white?" No: given the fact that what cats there are in my apartment are black and white is *accidentally* true, we cannot hope to rely on the property of being in my apartment to reliably, predictably, and necessarily to generate black and white cats.

The statement, "Sugar dissolves in liquid," on the other hand, does entail the counterfactual "If this were sugar, it would dissolve in liquid," and, as such, is thought to be an instance of a natural law. It is sufficient, that is, to bring it about that something will dissolve in liquid if it is so naturally disposed; and sugar is one of those of things with a disposition to dissolve.³

This distinction between accidentally true generalizations and those that are counterfactual-supporting is one that can help to distinguish those descriptions of events that have a role in the explanations involving those events, and those that do not. Some descriptions of events will generate laws by way of supporting counterfactuals and will thus be pertinent to explanations; and some will not, although many of those that are not can be used, of course, to identify events which are causes.

Now it is important to point out that it is possible to reject the causal criterion of scientific explanation (see Salmon et al., 1992). One might even reject the idea that laws of any kind are needed in order to get to

mechanisms, and so to explain how things work (Schiffer 1991). Davidson, of course, is famous for claiming that psychological explanation can't be got at by the positing of laws (1980). The issue of competing models of explanation is one that will get some attention below, as it bears on the question of whether an externalist criterion of content individuation is compatible with the causal relevance of content.⁴

To return to the issue of primary concern: we have seen that a causal statement will contain singular terms referring to the causes and to the effects. The question is: do the descriptions occurring in those statements have an explanatory role or not? Let us see how this question applies to the case of mental events, their causes and effects, and psychological explanation.

3. Mental causation

Presupposing that mental events can be causes, we can ask the question: what does this amount to? A believer in mental causation would presumably take it that there are true singular causal statements referring to mental events:

(3) Pain caused her teeth to clench.

(4) Her thinking it was lunch time, and her desiring lunch, caused her to ring for the parlourmaid.

The question now is whether mental descriptions ('pain,' 'thinking,' 'desiring,') are explanatory or merely accidental. That is, is it in virtue of having *mental* properties that these events have the effects they do have?

The question is particularly pressing since mental events are thought to have a physical realization, so it is certainly not at all obvious that it will be the mental descriptions that have an explanatory role. Yablo (1992, 248-249) expresses the worry:

. . . the mystery is how mental events, desires for example, can be making a causal difference when their unsupplemented neurophysiological underpinnings are already sufficient to the task at hand . . . mental events are effective, maybe, but not by way of their mental *properties*; any causal role that the latter might have hoped to play is occupied already by their physical rivals.

Of course, just because a mental event is described physically doesn't mean that the physical description will be ipso facto explanatorily pertinent: surely there can be many physical properties of an object or event that simply reveal nothing about the causal mechanisms at work. The point here is only that there are at least two descriptions of a mental event that can operate at any given time, so there is no logical reason to suppose that the mental description will be the one that will pick out the explanatory properties.

As we saw in the general case, causal relations are

referentially transparent. Causes can be picked out without its being specified which of their characteristics have an explanatory role; that is, a cause can be referred to without its being specified what it is about it that brings about its effect. Making a causal statement does not entail making an explanatory one. So, for instance, even if you denied epiphenomenalism for mental events, it wouldn't follow from that that descriptions of mental events tell you what you need to know to explain their effects. Is there any real reason not to abandon mental talk in explanatory contexts?

4. Belief-Desire Psychology

Given the above quite familiar points one obvious option is simply to deny that mental descriptions are causally explanatory. This of course may or may not go along with the further option of denying that mental states are causal engines in the first place.⁵ The problem with these options, however, is that they appear to fly in the face of a commonsense theory about psychological explanation.

Human behavior, as well as that of some animals, is commonly and plausibly thought to be the outcome of a certain amount of mental perturbation. When we want to know why A kissed B, it helps to be told that A believed B was her lover, and that A desired to show her lover some affection. When we want to know why the cat is yowling, it

helps to be told that he wants food. Behavior, that is, seems at least partly produced by mental states with content; and the content of those states appears to tell you much of what you need to know to explain behavior. So a discipline whose aim is to categorize, characterize and explain behavior seems to need to include mental states in its roster.⁶ Psychology plausibly proceeds by way of propositional attitudes.

Strong words have been uttered concerning the feasibility of abandoning talk of content in psychology (Fodor 1987, p.xii):

. . . if commonsense intentional psychology really were to collapse, that would be beyond comparison the greatest intellectual catastrophe in the history of our species; if we're wrong about the mind, then that's the wrongest we've ever been about anything . . . We'll be in deep, deep trouble if we have to give it up.

Of course, Fodor concedes that folk psychology bears something of a mere family resemblance to the "rigorous and explicit intentional psychology that is our scientific goal" (1987, p.xii). No matter. In the end, it is *intentional* psychology that is our scientific goal. In fact, what it is that a properly scientific psychology should be is an issue that emerges as more problematic than Fodor makes it appear, inasmuch as the problem of content's explanatory role begins to look somewhat stubborn, inviting creative solutions. We will consider the issue of what should count as proper

psychology below. For the moment, we can launch the investigation into the status of content's explanatory role as follows: content-citing mental descriptions seem, for the most part, to be successful tools for predicting and explaining behavior. And while it is the case - presupposing a causal criterion of explanation - that it is possible to merely assert that mental descriptions have no explanatory utility, in fact there is a theory of *content* that has the unfortunate consequence of making it difficult to see how it is that content can play an explanatory role.

5. The Case of Factives: An Analogy

A useful analogy, both with respect to the issue of the explanatory relevance of mental descriptions in itself, and with respect to the complications engendered by externalism and psychological explanation, is the case of factive expressions: descriptions of mental states that imply that certain facts obtain: 'knowing that *p*,' 'remembering that *p*,' 'regretting that *p*,' 'realizing that *p*,' and 'perceiving that *p*,' for example.

Consider whether

(5) A went to Fairway to buy arugula because he knows that Fairway sells arugula

cites an explanatorily relevant characteristic of the mental state involved in A's going to Fairway to buy arugula. (5) seems reasonable, on the face of it, as an explanation. But there is an objection, which goes as follows. The explanation offered, which cites A's *knowing* that Fairway is a place to get arugula, is misleading. Suppose A only *thinks* he knows that Fairway is a place to get arugula; suppose Fairway is a dressmakers instead. The fact that A is not, after all, in a state of knowledge turns out to be neither here nor there with respect to his Fairway-directed movements - what really matters is what A *believes* with respect to Fairway. From the fact that his belief is false it doesn't follow that he doesn't possess it, and it is the property of possessing that *belief* about Fairway - though false - that is causally implicated in his doings, and that it would be explanatorily useful to cite.

The point holds for cases, say, of remembering and even seeing. Both of these imply facts about the world which are out of the range of the subject's psychology, which, intuitively, is located inside his head. But it isn't what a subject actually remembers that plays a part his psychology: it's what he *thinks* he remembers. And it isn't what a subject sees that play a role in his psychology, but what it seems to him that he sees.⁷

What a subject knows, sees or remembers can be expressed in mental descriptions that imply something quite

external to the subject: the state of the world.

Intuitively, it seems difficult to see how the state of the world outside the subject could make any difference to the subject's psychology. And so the correct thing to say about such cases would appear to be this: of course you can use factive descriptions of mental states to make causal statements about the doings of a subject, but, in so doing, you would not necessarily be citing the characteristics of his mental states that play a causally explanatory role. And if, say, you wanted to specify the taxonomic principles by which your science of human agency proceeded to make its distinctions, factive descriptions of mental events would be ruled out, for these do not necessarily pick out the explanatory factors that are operative in the doings of a psychological subject.

6. Externalism and the role of content in explanation

This brings us at last to the consequences of externalism for psychological explanation. In fact, the basic intuitive problem has been anticipated - as intended - by the analogy with factive expressions. A theory of content that claims that some or all attitude attributions use content clauses which imply something independent of the subject's psychology - the environment, both particular and abstract, that surrounds her - will similarly have to face

the possibility that such content clauses fail to have a genuinely explanatory role.

That there are consequences for the role of content in explanation given externalism is, on the face of it, a familiar issue, concerning which there is a vast literature. It is necessary, however, to revise some of the details associated with the orthodox view. Most in need of rethinking is the tacit belief that the explanatory role problem for content is its failure to supervene; that content becomes problematic in propositional attitude psychology *because* it fails to supervene; so, complementarily, if content does not fail to supervene there is no problem. Of course, given the rather widespread tendency to take Twin Earth cases as definitive (even necessary) for an understanding and formulation of an externalistic view, this conception of the problem of content's explanatory role is not surprising. But there is reason to think it is inaccurate.

The distinction between weak and strong externalism, defended in chapter 1, raises a question as to just what is to be understood by the consequences of externalism for the role of content in psychological explanation. We must understand this issue now as dividing into separate but related questions: What are strong externalism's consequences for explanation? Are there consequences for psychological explanation given weak externalism as well as

strong externalism? And what, if any, are the logical connections between these problems?

7. Non-supervenience, extrinsicness,
and redundancy

We can begin by briefly considering the familiar consequences of strong externalism for content-based psychology. Two presuppositions - these are not completely uncontroversial in themselves - are relevant to this issue. One, as we have already noted, is the causal character of explanation. The other is that mental states have representational or semantic properties as well as intrinsic ones. The question arises: which features of propositional content have a genuinely explanatory role? The explanatory claims of a content-based psychology, of course, advert essentially to the semantic features of content, taking more or less the form:

(6) *C* did *a* because she believed that *p* and desired that *q*.

But is it the semantic features of content that have an explanatory role? The problem, briefly, is as follows. On one criterion, psychological explanation, like all scientific explanation, is causal explanation. The features of a cause that lead to an effect must be located where the causal interaction takes place, on pain of making a mystery

of the causal process: invoking action at a distance and other occult phenomena. Mental causal interaction must thus be understood to take place in the head, as it is most likely to suppose that the relevant machinery to effect causal processes is going to be located there. But this of course sits uneasily with the contention that content fails to supervene on what's in the head. The tension is clear: the causal powers of content must be in the head, but content itself is not; a puzzling result.

As we noted above, it is a familiar story that strong externalism has untoward consequences for content in explanation. What may be less familiar, however, is the claim - to be defended here - that externalism in its most general formulation has problematic consequences for content's explanatory role; that is, an explanatory role problem arises for content whether or not content supervenes on inner states. What we are after, ultimately, is a way to formulate the problem of content's explanatory role so that it is clear that any concept for which externalistic determination conditions hold will fall afoul of it.⁸

It is worth showing how the orthodox formulation of the problem of content's explanatory role can be absorbed into a more general one. Some kinds of content, as we know, just imply that a person is on TE. And this, of course, implies something about the person's environment, the world they're in, the facts that obtain. Content doesn't supervene

on the inner states that do the real explanatory work, so content doesn't do any real explanatory work.

One quite popular way of understanding this is to say: *wide* content varies in this way, and given that its variation doesn't seem to make any difference to the doings of the subject, then wide content has no explanatory role. What *does* have an explanatory role is *narrow* content. That is, given that twins could do the same thing whether or not they both believe that *p*, both me and my twin could yelp and run away at the sight of what we both call "water," since the respective substances referred to by that word in our environments scare us. So there must be something that is in common between us that explains why we do the same thing. And it isn't our belief that *p*, when it is taxonomized widely; she doesn't believe that *H₂O* is scary (and I don't believe that *XYZ* is scary). What we *do* share is a *narrowly* taxonomized belief that *p*; and it is this that has the explanatory role.

Logically speaking, of course, this is exactly the same as the claim that *knowing that p* is not what explains someone's doings; rather, what really explains someone's doings is, say, *narrow knowledge*, stipulated to mean *believing that p*. Of course in this case it is somewhat easier to see that there is something in common between someone who believes truly that *p* and someone who believes falsely that *p*: they both believe that *p*. The fact that one

belief is true and the other false is a - dispensable - wide fact about them. Nevertheless, it is thought that some such distinction can be made for belief content itself.⁹

Now the arguments in chapter 1 put forward the case that only natural kind and indexical content fails to supervene. This of course suggests that two kinds of content - no more - fail to sustain a viable role in the explanatory claims of propositional attitude psychology (which, coincidentally, suggests that narrow content is a limited, even if successful, solution to the problem).¹⁰ Can we thus dismiss the problem of content's failure to supervene as too limited to be of any real worry? Or is there a wider aspect to the problem lurking beneath the surface?

Let us scrutinize a canonical failure of supervenience problem to see whether there is more to this sort of example than meets the eye. My twin is molecule-for-molecule identical to me, but when she believes that her colleague is a humorless prig, her mental state is not the same as mine, since we are not in the belief-relation to the same items. So there is no content-involving psychological generalization that covers our (by hypothesis similar) behavior. My twin does not roll her eyes because she believes *my* colleague is a humorless prig.

On reflection, it seems clear that there is an obvious question as to what is the truly relevant factor that drives the standard type of non-supervenience problem. Content

fails to supervene because, given externalist taxonomic criteria, it is individuated in virtue of head-independent entities. But isn't it the case that this fact alone is sufficient to generate an explanatory role problem?¹¹ The essence of the non-supervenience problem really appears to devolve on the reason *why* content fails to supervene in some cases. And the answer is clear: externalistically determined content is *extrinsic* to the boundaries of the causal machinery of the subjects of mental states, making it difficult to see how it could play an explanatory role. The failure of supervenience problem, when we think it through, more accurately emerges as an instance of the widespread problem that arises whenever the individuation conditions of any concept are held to necessarily involve entities extrinsic to inner states, the states most plausibly thought to bear the causal burden in explanations of behavior.

These remarks suggest a very natural preliminary way to describe the explanatory role problem that arises for content from an externalist perspective: descriptions of mental states, on this view, are extrinsic - they refer to entities outside the boundaries of what can intuitively be claimed to play a role in the internal causal works - the psychology - of a subject.¹² Commonsense thinking about what matters to a subject's psychology centers on, among other things, what *appears* to her to be the case; how things *look*, what she *believes*. And if something extrinsic to or

independent of these states can vary without resulting in a variation in these states, then, intuitively, that something will not play an explanatory role with respect to the subject's psychology.

A few remarks on the notion of extrinsicness might be useful, although it must be said that beyond the obvious, it is not an easy notion to define. It can be understood by way of the distinction between internal and external relations, a topic that has enjoyed at least one period of philosophical vogue (see Moore 1922, and Ryle 1935, among others). It won't do to become too embroiled in this topic, however, as it raises deeply controversial metaphysical issues like essentialism, accidental properties, and the like, which it is preferable to avoid here.¹³

Typically, extrinsicness is understood as a spatial relation: two things are extrinsic to one another if they do not enter into each other's natures, are not part of one another. With respect to the issue of content in psychological explanation, however, it is best not to take the usual spatial understanding of this notion too literally. We can stipulate that an object *a* is extrinsic to a system *b* if *a* is not part of the causal powers of *b* - *a*'s causal powers aren't a part of those of *b*, say. So if *b* is a person and *a* is a denotatum of *b*'s terms, then, since *a* is out there in the head-independent environment, its causal powers aren't part of the causal powers of *b*. But if *a* is

not part of the causal powers of *b*, then there seems to be a *prima facie* problem as to how it can be the case that being in relation to *a* is explanatory with respect to the doings of *b*.¹⁴ A question will arise - to be treated below - as to how best to understand the extrinsicness of the denotata of predicate expressions.

We have alluded, above and in chapter 1, to the prominent position occupied by the orthodox understanding of the failure of supervenience problem. But, as we have seen, the non-supervenience problem apparently affects only two kinds of contentful states. And as long as the problem of content in explanation is thought to be the problem of content's failure to supervene, we must conclude that the problem of content in explanation is a very limited problem.

But of course we have reason to suppose that there is more to the problem of content in explanation from an externalist perspective than the problem of content's failure to supervene, since we now have reason to suppose there is more to externalism than the failure of content to supervene. As we have seen, what really lies behind the failure of supervenience problem is content's *extrinsicness*. Extrinsicness is a ubiquitous feature of all content individuated according to externalist criteria, whether strongly or weakly external. And if we pause here to ask the question what conceivably could be the operative idea behind the usual formulation of 'content's explanatory role

problem,' what we see, after a moment's thought, is that the only thing that could mean is that referring to content in explanation is irrelevant; content has no explanatorily relevant role to play in causal explanations of behavior.

We are now in a position to formulate a completely general explanatory role problem for content. If content is extrinsic - as externalism weak and strong demands - then it will be *redundant* in psychological explanation. Consider the following: suppose someone insisted, in her explanations of a companion's behavior, on referring to his (justified true) *belief*-states by using the expression 'knowledge.' What would we make of this? It is, of course, possible to refer to someone's beliefs using the term 'knowledge' or to refer to someone's mental states by using contents that bring in the environment, but those modes of expression are irrelevant in the sense of being redundant. If they were omitted, clearly nothing by way of explanatory force would be lost. And since, in psychology, what we are interested in is explanation, it appears safe to conclude that they can be omitted, since they appear to have no explanatory function.

A number of examples can be culled from the literature to illustrate what is known as the redundancy problem (hereafter TRP; see McCulloch 1988; Noonan 1991; Peacocke 1981; Segal 1989).¹⁵ Most of these remarks focus on the redundancy of singular thought content, but the basic point has been extended to general thought (McGinn 1989).

The best way to understand TRP is in terms of counterfactuals. As we noted in chapter 1, externalism in its most general formulation is best understood as a thesis concerning the possession conditions of mental states. The existence of extra-cranial objects and properties - the reference of the expressions in a psychological subject's language - is a necessary condition on the possession of those of her mental states involving those objects and properties. Should the relevant object or property fail to exist, no contentful state can be attributed.

But now consider the not implausible case where two people (or one person at different times) should perform the same behavior, and yet only one of them instantiate the relevant propositional attitude property. McCulloch gives an example¹⁶ (1988, 84-85):

Faith, Hope and Charity are as similar as you please but for some important details. They are physically very similar, are embedded in very similar environments, have very similar histories, and each seems to herself to be confronted by a dangerous cat. Each seems to think to herself a thought she would express thus: 'I'm going to kick that cat.' Each consequently lashes out. But here are the important differences. Faith and Hope confront distinct cats (Mildred and Consuela respectively), and Charity confronts no cat, but is hallucinating. Now some philosophers espouse the doctrine of Russellian thoughts (RT-doctrine), the doctrine that a thought standardly expressible by e.g. 'I'm going to kick that cat' essentially concerns the cat (if there is one) to which 'that cat' would refer. And they mean this to imply (a) that

Faith and Hope think different thoughts (since they think about different cats), and (b) that Charity thinks nothing of the corresponding type (since she confronts no cat) . . . the RT-doctrine is part of a claim that Intentional psychology as we normally conceive it is externalist, in the sense that seeings, believings and many other states of mind are taken in a relational or environment-involving way, as seeings or believings about this or that environmental item . . . Some reject the RT-doctrine. With respect to the above case, their essential reason would be that our subjects are so similar: subjectively, things seem the same to each of them; objectively, they each lash out in the same manner because of how things seem. So (the thought continues) the lashings ought to succumb to parallel psychological explanations. Yet the Russellian thought theorists seem to rule this out. According to them, there are psychological asymmetries among our subjects, and so if the actions are explained in the usual way by the ascription, . . . of thoughts entertained, the explanations will not be parallel.

The idea is this: the same behavior can, intuitively, ensue irrespective of the existence of an object or property, so citing an object-or-property-involving thought in accounting for the behavior is bound to be explanatorily irrelevant. It can't be that what explains Charity's behavior is her belief that that is a cat, since she possesses no such belief. Citing a cat-related belief in an attempt to explain the three women's behavior will be beside the point (two of them don't have the same thought because it's not about the same cat, and the other just doesn't have a comparable thought at all - though, by hypothesis, she

thinks she does), a violation of Ockham's maxim.¹⁷ Content in explanation is therefore causally redundant.

Another example, from Noonan, makes this explicit (1991, 3-4):

Suppose I kick a cat, and suppose my action is intentional under a description in which occurs a term denoting that cat . . . Imagine now a second situation in which, from my point of view, everything is the same, but in which, in fact, I am hallucinating a cat. Since this is so I presumably lash out at the cat I believe to be within kicking distance in exactly the same way as in the first situation. The difficulty for the opponent of [the Redundancy Thesis] is now to explain why I do so . . . [In the first situation] some of the object-dependent psychological states I was in, reference to which was essential to the explanation of my action there, are not present at all in the second, hallucinatory, situation. Moreover, my hallucination in the second situation does not make it possible to me to *think* anything I was not able to think in the first, veridical situation . . . if the contentful psychological states available to me in the hallucinatory situation suffice to explain my action there, no reference to psychological states with distinct contents will be required to explain my action in the veridical situation . . . no reference to any *object-dependent* content, not available in the hallucinatory situation, will be required to explain my action in the veridical situation, non-relationally described . . . For if the subset of contentful psychological states common to deluded HN and non-deluded HN suffices to explain the former's action, it must suffice to explain the latter's action, also, since the actions are identical - each, that is, *makes the same bodily movement*.

These examples ought to suffice to get the intuitive problem across.¹⁸ Now, the arguments in chapter 1 defended the role of part of the abstract environment, as well as the more immediate environment, in the determination of content. Properties, we saw there, are head-independent. And if being in a content-bearing mental state involves being in a relation to, among other things, a property, one is thereby and likewise in a relation to something quite extrinsic to the causal entities relevant both to effecting the transition from thought to action, and explanatorily pertinent.

Mathematics offers the most obvious preliminary example. Thoughts with, say, arithmetical content ("I hope that lifting a 20 pound weight for 3 sets of 12 repetitions each will tone the bicep") are expressed in sentences that contain expressions referring to numbers (number properties). Entertaining this thought puts one in the entertainment-relation to number properties, entities (pace anti-platonism) that are completely head-independent. Their extrinsicness puts them well beyond a plausible causal range with respect to affecting behavior.¹⁹

The formulation of the problem with respect to content that does supervene, however, is most clearly understood in terms of counterfactuals, as in the case of singular thought content. Again, the existence-and-identity conditions of general content are those of the reference of the terms of

the content-sentence: should these fail to exist, no content can be attributed. It is nevertheless plausible to suppose that a person could perform exactly the same behavior in the case where he does not genuinely possess the thought in question (but only appears to himself to so possess it²⁰) as in the case where he does.

A superficial objection can arise regarding this formulation of the problem; it is not a good objection, but it is instructive. It may be said that properties, as abstract objects, are *necessary* existents; so that the counterfactual premise regarding their content-determining and causally explanatory role can never obtain. But this is confused. In the first place, it is not obvious that conditionals with impossible antecedents can't be well-defined²¹: why isn't it perfectly coherent to say that if property *P* hadn't existed - even though it necessarily does - then its non-existence would make no difference to the behavior person (thinking himself to be) referring to that property. Although the counterfactual is harder to interpret for the case of properties, it is clear that the same *basis* for the redundancy problem of content applies.

Just to be sure this has penetrated, consider the parallel case concerning a state of knowledge. Take a limit case: a belief one can't help knowing (Descartes' *cogito*, say). Now surely, even for a case like this one, it is perfectly in order to raise the question whether knowledge

is an explanatory concept. Even in cases where a belief cannot be false, it is nevertheless the case that what is causally relevant, on the face of it, is the *belief*, not the state of knowledge. So a perfectly coherent question can arise regarding the explanatory role of knowledge, one that is mirrored in the formulation of the problem of weakly external content's explanatory role.²²

A related point, worth clearing up, concerns the metaphysics of properties. The claim that properties are necessary existents is sometimes defended on the ground that they are abstract objects. This would appear to result from uncritical dogma regarding the modal status of properties, a view subjected to criticism in chapter 1. We can add a counterexample to what we said there, so to put to rest any further difficulty concerning this point. Consider sets. Sets are abstract objects but their existence-and-identity conditions are those of their members; and the set of cats, it must be conceded, does not exist necessarily.²³

Putting the issue of the modal status of properties to one side, it is now clear that there is more to the problem of content in explanation than its mere failure to supervene. Even when content does supervene, externalistic criteria require that it be individuated with respect to head-extrinsic entities: properties. Two related problems thus ensue (both of which also apply to content that fails to supervene): extrinsic entities are not situated at the

causal nexus, so to speak, and so are explanatorily inert. And if we ask what difference the extrinsic, head-independent world makes to behavior, one quite plausible answer is none. Even when an object or property fails to exist, one can perfectly well be thought to behave in exactly the same way, in spite of the fact that one fails to be genuinely belief-related to an item out there in the world. To advert to the extrinsic objects of thoughts in explaining behavior is thus causally redundant.

A few clarificatory remarks are in order, especially with regard to the idea of a property - an abstract entity - being 'outside' or 'extrinsic to' the causal nexus. This extrinsicness certainly should not be thought of as a spatial issue. Confusion on this point might lead to the following objection: since properties are abstract, they can't be at the causal nexus (they can't be *anywhere*; the causal nexus is a specific location in space and time. So, if not being at the causal nexus were sufficient to rule something out as being explanatory, then all properties would be ruled out as explanatory, by *reductio*.²⁴

This objection arises, as we suggested, by a too-literal interpretation of 'at (or not at) the causal nexus.' 'Not at the causal nexus' shouldn't be thought to mean 'not existing at that place'; rather, 'outside the causal nexus' has to be understood, with respect to properties, as to do with not being instantiated. After all, it appears to be a

minimal necessary condition (although not sufficient) of the explanatory relevance of a property that the property be instantiated: no property not instantiated by the table is going to be relevant to the causal powers of the table (its 'causal nexus'). A property *P* not instantiated by this table will be outside its causal nexus, thus, in this sense, and not in the (really rather puzzling) sense of not 'existing there.'

Properties can both be instantiated by and be the objects of mental states (as we have argued); confusing these two characteristics will interfere with grasp of the basic picture here. There is a difference between a property's being a content-constituent and its being instantiated. The point (the *problem*) about content in explanation is that the properties which are (arguably) part of the content are not *instantiated* by the person, the subject of the mental states in question. And it would be an obviously desperate and unsuccessful move to try to save the causal relevance of some set of properties by claiming that although properties are abstract universals which don't exist *at* or *in* the person, nevertheless they are instantiated by the person. In general, of course, one does not instantiate every property one can think about. The property that *is* instantiated by a subject is the property of *having belief b about a property* which, of course, is not the same property as the property one has the belief about.

It is the relationality of content that results in content's explanatory role problem. What we have seen here, however, is that content is not just a relation to particulars, but also to properties via the predicate-expressions in content-bearing sentences. The properties that individuate content interfere with its causal role in that they themselves are not instantiated by the subject of the mental state in question. Properties fail a quite minimal condition on causal relevance - and long before any questions to do with their suitability for being the kinds of things that can be causally relevant can be raised.²⁵

Formulating content's explanatory role problem in terms of redundancy is useful in that it undermines what appears to be the tacit understanding that there are no problematic consequences for content unless it fails to supervene. What we now see is that all concepts to which externalistic determination conditions apply are extrinsic to the causal machinery to which the fundamentals of mechanical engineering - even mental engineering - must apply.

8. Non-supervenience, Extrinsicness and Redundancy:

Entailment Relations

It is worthwhile to pause here to spell out the entailment relations between the problems with which we have been concerned. First: that content is extrinsic doesn't

entail that it fails to supervene: you can't run a TE case for mathematical content (as we saw in chapter 1), but, as explanations involving mathematical content will advert to entities quite extrinsic to the head, there remains the problem of adverting to content that is not situated in the requisite causal location. It seems plausible to suppose, however, that when content fails to supervene, it does so because of its determination in virtue of entities extrinsic to the head.²⁶

Similarly, it will not strictly speaking follow from the fact that content is extrinsically determined that it will be explanatorily redundant. There are models of explanation, in fact, where being extrinsic to the head is a positive boon: a functional/teleological criterion of explanation is one such. The redundancy of content in explanation will follow, however, on the presupposition that explanation is meant to be causal. Extrinsic entities - particular or abstract - are simply not in the right location to play the required explanatory role, if what is relevant to explanation are the causal processes involved.

Finally, content might be redundant in explanation without it following from that it is externalistically individuated, of course. That *x* is redundant in a science, in a theory, or in an explanation is an extremely general complaint to make of it. *X* could be redundant for any number of reasons: it might be magical, meaningless, unverified,

platonic, intrinsic. The terms of the inquiry set the limits of relevance.

9. The Redundancy Problem: Distinctions

There are connections between the redundancy problem as we have formulated it and two other issues that can arise with respect to the problem of content in explanation that also involve a notion of redundancy. The issues we are about to consider are by no means equivalent, but they do have enough points of contact to cause confusion, and to make it imperative to delineate very carefully between them.

a. Syntactic Properties vs. Semantic Properties in Explanation.

There is a line of thought, associated with Stich (1978, 1983) and Field (1978), that urges the repudiation of semantic properties of content for the purposes of psychological explanation.²⁷ Stich, for example, writes (1983, 8-9):

The core of my argument is the claim that the theorist who seeks to couch the generalizations of cognitive science in the content-ascribing language of folk psychology will be plagued by problems that can readily be avoided . . . I shall sketch an alternative paradigm for cognitive theories which avoids the problems engendered by appeals to content . . . The alternative is what I will call the *Syntactic Theory of Mind*. Cognitive theories which cleave to the STM pattern treat mental states as relations to purely syntactic mental

sentence tokens, and they detail the interactions among mental states in terms of the formal or syntactic properties of these tokens.

According to this view, it is the syntactic properties of mental states that are the explanatorily relevant ones, and Stich, for one, gives arguments are meant to show that "folk locutions are scientifically otiose" (1983, 9). Stich's position with respect to content in explanation is, in effect, indeed one that charges content with irrelevance given the demands of a properly erected cognitive science. But, as we said above, redundancy is an extremely general notion. What externalism contributes with respect to explanation is indeed also the idea that content is redundant in psychological explanation: the question is why. The answer will distinguish TRP from a position like that of Stich.

As we noted above, what lies behind TRP is a counterfactual idea: given the individuation-conditions on content imposed by externalism, and the parallel content possession-conditions that result, it is clear that no mental state can be ascribed if the object of the mental state (a particular, in the case of singular thoughts; a property, in the case of general thoughts) does not exist. But two subjects (or one subject at different times) can behave exactly the same way, in spite of the fact that only one of them can be ascribed a mental state with some content C. The explanatory generalization we desire will thus have

to proceed other than by citing the content of their thoughts.²⁸ On the other hand, a view that takes the syntactic properties of mental states to be all that need figure in properly scientific explanations can obviously be defended quite independently of anything to do with semantic externalism.²⁹

To see this, consider the familiar *Language of Thought* hypothesis.³⁰ Whenever a belief is ascribed, it is surmised that there is a sentence in Mentalese that corresponds to the belief, and believing that *p* is formulated as a relation to that Mentalese sentence. Mentalese sentences, being sentences in the language of thought, are, of course, in the head. There are, of course, various effects of possessing the belief in question, which are what are meant to be explained by citing the belief.

Now, note that the Language of Thought hypothesis can be in principle paired with any number of theses concerning content and its individuation. But given that, on this view, to ascribe a belief that *p* is always to ascribe a corresponding sentence *s* of Mentalese, clearly one could consider explaining another's behavior by citing the Mentalese sentence; instead of (roughly): 'when *C* believes that *p* she does *a*,' rather: 'when *C* tokens sentence *s* she does *a*.' And it appears reasonable to propose that should it be the case that the syntactic properties of the mental state prove to capture all and more of the generalizations

we require than do the semantic ones, that we jettison the semantic ones as excess baggage. It is worth pointing out that this position is more radical than the one that favors narrow content over broad content in explanation: for, after all, a notion of narrow content is still a semantic notion. If sentences in the head have both shape and narrow content, and their shape properties do succeed in doing the required explanatory work, it is reasonable to ask what point there is in speaking of these sentences as having truth-conditions at all - even when those truth-conditions are narrow ones.³¹

In short, the kind of view that defends the explanatory role of syntax as being more adequate than that of content results in the idea that semantic properties are redundant in explanation, but does so in a completely different way, and according to quite different criteria, than does TRP. What Stich calls STM does not, for instance, make a counterfactual point in claiming that even narrow states have syntactic properties, so why bother citing the narrow states when the syntactic properties do the explanatory job; the point is not that even without narrow content the syntactic properties of a mental state are still on the scene. Rather, those that defend the explanatory superiority of syntax do so on the grounds that there is a kind of overdetermination problem that results when we compare the explanatory utility of syntactic and semantic

properties; one that, as it happens, can be formulated as an invitation to apply Ockham's razor in favor of syntax.

The notion of overdetermination ushers in another kind of redundancy claim that must be distinguished from the one that underlies our formulation of TRP. Though somewhat controversial, there is a view that the sciences form a hierarchy, from the 'basic' (physics) to the 'non-basic' (sociology or anthropology), roughly speaking. The non-basic sciences are sometimes regarded as being explanatorily useless, or redundant. The question is, on the basis of what is the redundancy of the special sciences thought to obtain? Should special science redundancy be equivalent to that which results for externalistically individuated content in psychological explanation, then the latter is a) a tacit repetition of familiar points, and so uninteresting; and b) clearly vulnerable, on that score, to a host of telling objections to special science redundancy.

b. Redundancy and the special sciences.

What we are concerned to point out here is, in fact, the disanalogy between the causal redundancy of content engendered by externalism, and the familiar - though contested - point that special sciences (psychology, for instance) are explanatorily redundant. The problem of content in explanation as we have been formulating it is not to be confused with the objection to special sciences, and

one of the ways to make the disanalogy explicit is to focus on whether the scepticism regarding content's explanatory role arises within the framework of objections to the explanatory relevance of special sciences in general, or whether, as we have been suggesting, the consequences of externalism for content's explanatory role are specific to *externalism*.

The explanatory relation between the special sciences and basic science is an issue that emerges from the wider one of reductionism in the sciences. Hempel comments (1969, 179):

. . . one further reason for the fascination the subject [of reduction] has held for philosophers lies, I think, in the ontological roots of many questions concerning reduction - questions such as these: Are mental states nothing else but brain states? Are social phenomena simply compounds of individual modes of behavior? Are living organisms no more than complex physicochemical systems? Are the objects of our everyday experience nothing else than swarms of electrons and other subatomic particles? Or is it the case, as the doctrine of emergence would have it, that as we move from subatomic particles to atoms and molecules, to macroscopic objects, to living organisms, to individual human minds, and to social and cultural phenomena, we encounter at each stage various novel phenomena which are irreducible, which cannot be accounted for in terms of anything that is to be found on the preceding levels?

It is sometimes thought that the sciences are arranged in a hierarchy, according to a certain criterion of property

instantiation. Not all objects in the world instantiate the same sorts of properties, so, plausibly, a particular theory will quantify over properties that another does not. So, for instance: all physical objects in the world have physical properties, but not all of them have chemical properties - some don't form molecules. Of those that form molecules, not all have biochemical properties (not all are living things). Now take all the things with biochemical properties - do this set belong to the set of things with biological properties? Putting aside a detailed consideration of the question of what counts as a biological property, suppose you stipulated that a biological property is a functional property. Then not every piece of biochemical material has a function, clearly - some of these things could be, for instance, small freestanding bits of tissue. Finally, not everything that instantiates biological properties instantiates psychological properties.

Now consider an account of an object in the vocabulary of a theory of, say, evolutionary biology. Say that object is a person. This is an object with psychological properties, and it is true that all the objects with psychological properties that we know of have biological properties, and those that satisfy *biological* descriptions satisfy *biochemical* descriptions . . . and so on down. But when you are talking about an object from the point of view of evolutionary biology, and in those terms, it is plausible

to claim that what you are not doing is talking, say, *biochemistry*; you are, presumably, at a higher level of generality. The controversy surrounds the question whether it follows that a higher level of generality means an explanation of little or no utility.

The methodological presumption behind reductionism in the sciences is that sound scientific generalizations are found only at the level of physics and possibly chemistry. Theories of any other kind - and their explanations - are thought to have to come under the rubric of physics in order to be sanctioned as making properly scientific contributions. This absorption is to be effected by certain principles of theory reduction. Nagel's work (1947, 1961) is the *locus classicus* for this notion. Briefly, reduction applies intertheoretically, and is accomplished by interpreting the non-logical vocabulary of the target theory into that of the so-called reducing theory. Combining this idea with the conditions imposed by the deductive-nomological model of explanation results in the further idea that the explanations of the now-subsumed theory can be derived from premises in the vocabulary of the reducing theory.³²

An enthusiastic defense of reduction with respect to folk psychology has been recently given by Churchland (1986, 395-396; 399):

Once folk psychology is held at arm's length and evaluated for theoretical

strength in the way that any theory is evaluated, the more folkishly inept, soft, and narrow it seems to be Insofar as it is a theory, it is an empirical, not an a priori question how good a theory in fact it is, and its 'obviousness' will not protect it from revision or replacement if it is flawed . . . it would be astonishing if folk-psychology, alone among folk theories, was essentially correct. The mind-brain is exceedingly complex, and it seems unlikely that primitive folk would have lit upon the correct theoretical framework to explain its nature Mental states may be functional states, but this does not imply that the specification of their functional profile based in folk psychology is correct Nor does it imply that psychology cannot be reduced to neuroscience. The claims for the autonomy of psychology are therefore misbegotten.

This is a rough approximation of the line of thought: it is, in fact, far from being universally accepted - particularly with respect to the elimination of folk psychology.³³ Now there are obviously many questions that arise concerning the relations between the putative levels of scientific theories, their laws and their explanations, and the autonomy of psychology, all of which must be set aside here. I wish to focus instead on a more fundamental question: on what basis are the special sciences thought to be explanatorily redundant? Remember that our main concern is to distinguish the core of the objection to the explanatory relevance of the special sciences from that of the objection to the explanatory relevance of extrinsic content in psychological explanation. And this is simply

stated. The special sciences are thought to be explanatorily irrelevant because, the objection to them goes, you won't lose any explanatory force by failing to cite them. No explanatory force is lost because the underlying or counterpart laws of *physics* - to which all these other vocabularies will *reduce* - capture all the generalities you need.

Now this is a highly general - if disputable - claim, and it applies to, say, chemistry, just as much to psychology. Nothing in this line of argument against the special sciences depends on the extrinsic individuation of content. Nobody who objects to the special sciences in terms of their explanatory utility is concerned even remotely with the extrinsic individuation of content, I'll wager. The argument concerning the redundancy of the special sciences arises quite independent of any issues to do with the individuation of content: it can arise if you are an *internalist*; it can arise if you didn't even believe in psychology at all (it applies to biology, to chemistry, to whatever *isn't physics*).

It is particularly important to avoid assimilating content's explanatory role problem (as we have been considering it) with objections to special sciences, for this may easily result in the error of supposing that the objections to content's explanatory role can be overcome by countering the objections to special science explanations.

But an example will show that such a move won't work.

Someone who thinks that objections to content in explanation stem from a failure to properly understand the explanatory relation of the special sciences to basic science, might, roughly speaking, make the following point, alluded to above: there is no reason to suppose that the move to a higher level of theoretical generality (assuming it even makes sense to think of the sciences as arranged in a hierarchy) entails, in some sense, a less effective explanation. If what we are interested in is, say, whether it is the fact that Elizabeth believes that Darcy is ungentleman-like that leads her to refuse him, and what kind of generalizations can be constructed as a result, then the laws of, say, gravity - though presumably applicable - won't help. Why can't there be, that is, *more than one explanation of the same thing*. Imagine, analogously, the state of pornographic literature if all we could give were descriptions and explanations couched wholly in the vocabulary of basic physics; of course these terms apply - but that's not what we want from our dirty magazines.

The multiplicity of explanations point is, arguably, quite effective in the special vs. basic science debate, but it is not terribly convincing when it is applied to the problem of extrinsic content's explanatory role. Consider the familiar formulation (now, of course, superseded) of the problem in terms of content's failure to supervene: it does

not seem very effective to counter the argument with the claim that there is every reason to suppose that there can be more than one kind of explanation for the same thing. The failure of supervenience problem, as we have seen, depends on certain quite specific issues raised by, in part, the nature of *causation*. The explanatory relation between special science and basic science is quite orthogonal to this.

Consider, to drive this home, what someone who was confused on just this sort of point might say in objection to narrow content. "Jerry," they'd likely sniff, "when you say that broad content isn't explanatory you're just forgetting that there are special sciences, and in one/some of them you've got broad content. The fact that there's another science that has narrow content and that explains the same events is neither here nor there. There can be more than one explanation of the same thing, you know." The right reply to this, of course, would be to say, "but there's a *special* reason why that view - one that in general has something to recommend it - doesn't apply here: and this special reason has to do with failures of supervenience. In those *other* special sciences the chemical, say, *does* supervene on the physical, so there's no violation. But a whole different problem emerges if broad content doesn't supervene on the physical. . . ."

On reflection, we can make the point that even if

psychology were basic, one could still raise the question whether all of its descriptions were causally relevant - after all, in physics, I can describe a collision of particles using a causally irrelevant description. It would be, nevertheless, a description of elementary particles, but one with little or no explanatory role. And, ultimately, even if Cartesian immaterialism were true - even if there were *no* supervenience on the physical - one could still at least *raise* the question which mental descriptions are the causally relevant ones. Perhaps some are, perhaps some aren't, perhaps all are, etc. The answer to the question would have to proceed along the usual lines; the point is, the question is not empty, and none of these points have anything to do with the issue of special science redundancy.

In sum: it is confused to suppose that a hierarchy of the sciences and the availability of subsequent counterpart basic level explanations are sufficient reasons to entail an elimination of mental terminology from explanations, and further confused to suppose that any and all objections to mental terminology in explanations are just versions of that point. Redundancy is a very general concept, and more or less always means the same thing: what distinguishes TRP as we are considering it from both syntactic views about the mind and special science redundancy is what *motivates* TRP.³⁴ The motivations have been enumerated above. It now should be quite clear that TRP cannot be said to presuppose

any view at all about the status of the special science with respect to explanation, and, as a result, it should be clear that TRP cannot be overcome by way of counter-objections to scepticism about the explanatory relevance of the special sciences.

10. Resolving the problem of content's explanatory role:
the options

We have clarified and cleared aside a number of points regarding the formulation of extrinsic content's explanatory role problem. What concerns us now is to consider the options for a solution.

Let us recall, once more, the core of the problem.

McGinn comments (1989, 133):

Basically the point could be stated thus: the reference relation, as between symbol and object, does not contribute to the causal powers of the symbol - it is not what empowers the symbol to bring about its effects. The causal mechanism whereby the symbol has an impact on the world does not somehow incorporate the relation of reference. The relation of reference is to the symbol what the country of origin of a car is to its engine, i.e. not part of the causal machinery. Intentionality is not what makes the world go from one state to the next. Content is not a *mechanistic* feature of the world.

Different authors³⁵ put this in different ways: mental causes must be supervenient on internal states of the subject; the only thing with a causal role is the "shape" of a mental symbol; the characterizations of causal mental

processes must respect a "formality constraint"; the cognitive operating system is causally a "syntactic engine"; mental algorithms can be causally sensitive only to local features of their inputs; truth conditions cannot play any role in causal-role or functional psychology; it is a sort of category mistake to attribute causal potential to meanings; a causal-explanatory taxonomy of mental states must be narrowly individuated. In spite of the fact that all of these formulations fail to recognize a truly widespread problem with respect to content - taking it for granted that failure of supervenience is a necessary condition for the problem - nevertheless, all attempt to accommodate what appears to be an undeniable: mental explanations must cite local, proximate, intrinsic causes. So the causal powers of a mental state cannot depend essentially on relations to entities whose existence is extrinsic to the body. But externalism says that contentful states are identified by just those entities. So any science devoted to uncovering the causal laws and mechanisms of the mind cannot do so by reference to contentful states; or, at the very least, any science that purports both to refer to contentful states and to uncover causal mechanisms will suffer the usual instability of any mismatched alliance.

So, in some sense, the issue really begins here. What then are the alternatives for a solution to this problem? We can preliminarily, and quite abstractly, characterize the

main contenders as follows:

1. Externalism is not at all inconsistent with causal relevance.
2. It is indeed inconsistent; what is necessary is a notion of narrow content.
3. No notion of narrow content can really be defined, and yet the argument against wide content is correct, therefore, eliminativism about content is the only solution.
4. It is not necessarily the case that explanatory taxonomy need proceed along causal lines; there are other, perfectly respectable sciences that individuate according to other criteria.³⁶

There is also a view - which has both a radical and a moderate side - which, although not unfamiliar, seems to have been somewhat neglected in the most recent debate. I believe it is interesting enough to resurrect here. This is the view that belief-desire psychology is not at all a form of causal theory; rather, it is a radically different kind of understanding, involving a notion of *rationalization*. While it is true that this view and its associated repercussions has enjoyed some vogue in the past, what is worth considering now is the intersection between it and the issues raised by *externalism* and explanation. What seems to

be the case is that such a view has the effect of undercutting the arguments against wide content, as psychology is thought not to be in the business of giving *causal* explanations - certainly something at which it is worth taking a closer look.

All of these views have different consequences for the issue as to whether content is something to preserve or reject in psychological explanation. And the key issue with respect to which this issue should be understood is that of the relation between individuation and causation. Wide content is thought to have to go when the dominant criterion is that of causation, with its concomitant imposition of *locality*, with which individuation is supposed to fall in, under threat of emasculation. A view that challenges that criterion, unsurprisingly, will have very different results for content in explanation.

So let us consider the alternatives for a solution to the problem of content in explanation according to how they view the individuation/causation connection.

a. The consistency of explanatory relevance and externalism.

We can begin with the the claim that externalism poses no problem for the explanatory relevance of content. This is a view associated with McDowell (1980;1984;1986), Evans (1982), and Peacocke (1981, *inter alia*). The most effective attack comes from, among others, Fodor (1988); McGinn

(1991); Noonan (1986); and Segal (1989).

The view takes it that individuation - at least of singular content - is relational, and no explanation of an action directed onto a particular object can proceed other than by citing such content. So not only is wide content explanatorily relevant, it is downright *indispensable* in explanations of certain kinds of behavior. Peacocke, for instance, claims, (1981, 205):

. . . No set of attitudes gives a satisfactory psychological explanation of a person's acting on a given object unless the content of those attitudes includes a demonstrative mode of presentation of that object.

This view is, of course, illustrated by way of *demonstrative* content - content subject, as we have seen, to *strongly* external individuation. In spite of the usual difficulties attendant on the suitability of wide content in explanation, Peacocke sees no problem, in accounting for the explanatory relevance of such content (1981, 198-199):

. . . the obvious reply is that in the case of your belief that pen is valuable, your belief and the corresponding belief of your *Doppelganger* are not psychologically indistinguishable, because your belief causes you to act on the pen in front of you and *his* belief causes him to act on the pen in front of him. These are distinct pens.

That is: different objects of thought, so different causal powers of thoughts; different causal powers, so different thoughts. Wide content has an indispensable

explanatory role, at least with respect to actions directed onto particular demonstrative objects. Having a 'pen' thought has different causal powers than having a 'twinpen' thought; in fact, according to this argument, a 'pen' thought is a different causal power than a 'twinpen' thought. These are distinct pens.

Both Fodor and McGinn make the point that this is just confused. Fodor (1988, 34-35):

. . . so, to summarize, if you're interested in causal explanation, it would be mad to distinguish between Oscar's brain states and Oscar2's; their mental states have identical causal powers . . . it's true that I say 'water' I get water and when my Twin says 'water' he gets XYZ. But that's irrelevant to the question about identity of causal powers . . . what is relevant to the question of identity of causal powers is the following pair of counterfactuals: (a) If his utterance (/thought) had occurred in my context, it *would have had* the effects that my utterance (/thought) did have; and (b) if my utterance (/thought) had occurred in his context, it *would have had* the effects that his utterance (/thought) did have.

McGinn concurs (1991, 585):

The obvious fact that behaviour admits of relational description in terms of environmental entities has, we know, been triumphantly seized upon by opponents of the kind of causal thesis defended here: these opponents think it shows that there must be a difference between the causal powers of my beliefs and my twin's - after all, I reach for a drink of water (H₂O) and he reaches for retaw (XYZ)! Fodor (1991) definitively puts this 'argument' out of its misery (if it has not already expired of

natural causes). You might as well say that my knife and its molecular double have different causal powers because they each cut numerically different loaves of bread - or do it at different times and places!

Now what about cases where subjects are *deluded* in their, say, demonstrative perceptions? We have already seen that such subjects can perfectly well be thought to *behave* in exactly the same way as their non-deluded counterparts - but no thought content can be brought to bear in an explanation of such behavior - since there is no such content possessed. In fact, as Segal (1989, 43) points out, this kind of view (the Russellian or singular thought theory) can't make sense of "actions of subjects of empty singular thoughts . . . without undermining itself." In fact, what he does is press the redundancy objection the same way that Noonan (1991), does above. Segal concludes (p.45):

[singular thought theory] faces an unpleasant dilemma: either object-dependent thoughts . . . are always explanatorily redundant, or sometimes the actions of subjects who act on the basis of what they take to be singular thoughts are not rationally explicable. But neither of these options is acceptable. If we can explain all of a given subject's actions under intentional descriptions without attributing object-dependent thoughts, then surely we have no basis at all for making such attributions.

In sum, according to this kind of view, both relational individuation and explanatory relevance can apply

to content - but only because, as we have seen, the notion of causal power or causal role has not been properly understood. However misunderstood, there is a general effort on the part of this kind of view to make causation the criterion of individuation: the mistake is to characterize that criterion, in our terms, extrinsically.

b. Syntax

Earlier (section 8, (a)), we considered the view that folk psychological categories - beliefs, desires and the like - have been superseded, and that, on pain of irrelevance, they ought to be excised from any science of the mind. Here we will limit ourselves to a few comments.

First: how does the syntacticalist interpret the individuation/causation relation? Such a view that takes the goal of specifying the causal mechanisms underlying behavior to be the only one that a properly scientific enterprise ought to be concerned with. Stich notes (1978, 575-576, emphasis his):

. . . the principle of psychological autonomy states that the properties and relations to be invoked in an explanatory psychological theory must be supervenient upon the *current, internal, physical* properties and relations of organisms (i.e. just those properties that an organism shares with all of its replicas) . . . in specifying that only internal properties and relations are relevant to explanatory psychological properties, the autonomy principle decrees that relations between an organism and its external environment

are irrelevant to the organism's current (explanatory) psychological state.

Representational or semantic properties of content, of course, (sometimes) do not supervene, so cannot bear the required mechanistic role. Stich is explicit in his understanding of what the individuation/causation relation means for belief/desire explanation (1978, 578):

I will argue that if the autonomy principle is accepted then there are large numbers of belief properties that cannot play a role in an explanatory psychological theory . . . I think the belief/desire thesis can be profitably viewed as the speculation that these intuitively sanctioned singular causal statements can be cashed out in a serious psychological theory couched in terms of beliefs and desires. In showing that large numbers of these singular causal statements cannot be cashed out in this way, we make the speculation appear idle and unmotivated.³⁷

What can bear the required causal role are cognitive states (1983, 149):

whose interaction is (in part) responsible for behavior . . . systematically mapped to abstract syntactic objects in such a way that causal interactions among cognitive states, as well as causal links with stimuli and behavioral events, can be described in terms of the syntactic properties and relations of the abstract objects to which the cognitive states are mapped . . . the idea is that causal relations among cognitive states mirror formal relations among syntactic objects.

Differently expressed but essentially consistent is what Field (1981, 100-101) proposes: a system of internal

representation, or a narrow psychology:

. . . the syntax and type-identity conditions for a system of internal representation should be regarded as functionally characterized by a psychological theory in which they appear; and we can take that theory to be narrow psychology, that is, the kind of psychology that does not employ any semantic characterizations of the sentences in a system of representation. This is important, for it means that the syntax and conditions of type-identity for the system of representation could in principle be determined independently of any considerations about what the sentences in the system mean.

This sort of view is explicit about the individuation/causation question. Causal considerations act as the primary constraint on individuation on this view, since they act as the primary constraint on any properly scientific enterprise. And semantic properties - since they fail to supervene - are not such as to be caught in the causal net.³⁸

c. Narrow Content

The notion of narrow content, first proposed by Fodor (1988) as a solution to the problem of content in explanation, is excessively familiar; and as in our remarks above we have in effect provided some essential evaluative points, we can here be brief.³⁹

How does Fodor interpret the individuation/causation connection? He notes (1988, 30-32):

. . . scientific psychological

explanation, like commonsense belief/desire explanation, is committed to states to which semantic and causal properties are simultaneously ascribable . . . and I think it's quite easy to see how the required principles of individuation should be formulated . . . individuation in science is always *individualistic*. Common sense postulates a relational taxonomy for the attitudes, psychology postulates states that have content but are individualistic, so the question arises what notion of content survives this shift in criteria of individuation.

The scientific theory that vindicates belief-desire psychology is the computational conception of mental processes. The mind is conceived of as a symbol manipulating system, and mental symbols have both semantic and causal properties, which Fodor believes is the only way even to formulate an - or the - important question about the mind, namely, how it is that its causal processes make sense. But wide content falls afoul of a constraint on scientific theories: these, according to Fodor, are in the business of giving causal explanations ("of course," p. 34), and wide content cannot have a causal role because it doesn't supervene on the proximate bearers of the causal properties presumably involved in the production of behavior - internal, often physical, states. Enter narrow content: a notion of content that "survives the shift in criteria" since it cannot be sundered from its causal powers.

Fodor is thus crystal clear about the individuation/causation question: these stand or fall

together, and, on pain of postulating causal mechanisms the likes of which the world has never seen, they stand united internally to the subject organism.

The notion of narrow content has come in for its share of criticism,⁴⁰ and as my intention here is to consider the alternatives for a solution to the problem of content in explanation from a particular perspective, rather than to adjudicate these options, I will have little to add.

A few critical comments will, however, provide a natural segue to the next stage of these remarks. Recall that the interesting question insofar as these options are concerned is a common thread that runs through all but one of them: that is, how the relation between individuation and causation is to be characterized. As we shall shortly see, there are yet other options that reject the connection in question, with what success we shall see.

Confidence in narrow content as a solution to the explanatory role problem is eroded, as I suggested above, by the fact that it is only useful - if at all - in two cases: with respect to natural kind and indexical content, the only two kinds of content that fail to supervene. In cases where content does supervene, as we have seen, the extrinsicness of content generates the redundancy problem, a problem I have argued really subsumes the failure of supervenience problem, and for which narrow content fails to provide a relevant - even coherent - solution.

The problem for content in explanation, once again, arises when classification criteria for content don't coincide with what has to be the case about causality and causal powers. And whenever content is extrinsically classified, this problem will arise - the point is, content can be so classified, and yet supervene. The mistake is to think that the only time content gets pulled apart from its causal powers is when it fails to supervene. Isn't it just grotesque to apply the notion of narrow content to cases where content did not fail to supervene (even if it weren't the case that it is defined in terms of non-supervenience)? Consider, again, formal or mathematical concepts; we argued in chapter 1 that it was just not possible to coherently consider these cases where the internal and external properties could be independently varied, what needs to be the case to get the failure of supervenience off the ground. Let the sceptic imagine trying to propose a notion of narrow number, or narrow truth-function, to provide explanatory generalizations in cases where, say, my twin means something different from me when she says "not," but what she means seems the same to her as it does to me.

Further, should someone confuse the wide/narrow categories with the distinction between weak and strong externalism, one will embroil himself in no end of further confusion. Among other things, the status of the problem of content's explanatory role may appear more accounted for

than we have reason to suppose it is. Weakly external content is *not* equivalent to wide content: wide content is what varies in twin earth cases; what fails to supervene. And, as we spelled out in chapter 1, though it is the case that if it's wide it's weak, so to speak, it is not the case that if it's weak it's wide. We might - if we could be sure not to risk total referential chaos - stipulate a notion of *wide** content: *wide** content is to be understood in the sense of broad, global, extensive, sweeping - comprising, as we have argued, the particular and the abstract environment. When, however, wide content means what it is usually thought to mean, we would do well to remember that narrow content - at best - leaves much left over to be explained.

A final pedantic note. Fodor (1988, 27) says:

I am about to tell you two stories that you've very probably heard before. Having once told you the stories, I will then spend most of this chapter trying to puzzle out what, if anything, they have to do either with commonsense belief/desire explanation . . . The conclusion will be: not much . . . Indeed, contrary to the conclusion that I am driving toward, it is widely held that one or both stories have morals that tend to undermine the notion of content and thereby raise problems for propositional-attitude-based theories of mind.

The two stories are, of course, the Putnam/Burge twin earth cases, which, as we have seen, Fodor mistakenly takes to apply to all content. Be that as it may: another question arises. Contrary to Fodor's claim, it does seem as if

externalism genuinely raises a problem for propositional-attitude-based theories of mind - even if, like Fodor, you take externalism to be defined in terms of twin earth/failure of supervenience cases. Isn't the dialectical progression in Fodor's argument just that since the ordinary kind of content at work in ordinary commonsense psychology is too wide to accommodate the causal constraints on any scientific explanation, that we need this other - narrow - aspect of content? The point is that the ordinary commonsense content at work in folk psychology *isn't* narrow content. Fodor's claim that folk psychology can be saved from elimination by narrow content seems somewhat optimistic.

In any event, it may be that the very thing that Fodor takes as unarguable - namely, the causal nature of scientific explanation - is a constraint that ought to be relaxed in order to account for content's explanatory role. Clearly, as we have seen, all of the accounts that consider the criterion of individuation to be that of causation are faced with the problem of accommodating content's extrinsicness in a causal apparatus. It is unclear that any are terribly successful at resolving the problem. Let us consider two other accounts that distinguish individuation from causation and see whether they fare any better.⁴¹

d. Non-causal criteria of individuation

When Fodor sets his agenda for the issue of content in explanation, he - as we saw above - meets with incredulity the idea that scientific explanatory taxonomy could be anything but causal. Some views regarding the problem of content in explanation, however, are attempts to recalibrate the individuation/causation relation in just the way that Fodor finds unthinkable. It seems fair to say that the view associated with Burge (1986) tries to make causal *locality* compatible with *distal* individuation, a somewhat more radical claim than views associated with Millikan (1984). In urging an altogether non-causal criterion of individuation in explanation, the latter do not have to overcome the *prima facie* incompatibility of intrinsic causation and extrinsic individuation, whatever the rest of their difficulties. Dretske's (1981, 1988) views, are also, I believe, reasonably represented here among those who claim that non-causal criteria of explanation can make more profitable headway with content's explanatory role problem.

The sort of view in question might be generally summarized as follows. While it is true that referring to content in what purports to be a causal explanation will be problematical, to infer from that it need be eliminated from psychology is to suppose that the only legitimate way to individuate for the purpose of explanation in science is causally. However, it is contended, this doesn't follow. Why can't it be supposed that it is legitimate to individuate

entities in ways that are, in fact, not relevant to their causal powers. Among the sciences there are some - perfectly respectable - in which entities are individuated in ways that are not causally dictated. The mere fact that content and causation are not compatible is no reason to eliminate content from explanation - the solution is to individuate content non-causally.

Burge (1986) gives a sustained argument in defense of the thesis that individualism - understood as a view about how kinds are correctly individuated - is false for psychology. I will limit myself to comment on Burge's 'metaphysical' remarks, as it is these that most directly address the issue of concern here.

Individualism - the view that what is within the physical confines of an individual is all that can be relevant to, among other things, his psychology - seems to be a natural result of a causal constraint on individuation. The individuation of kinds in psychology have to be causal, so they have to be individualistic; there can be no other way to accommodate the causal requirement presupposed as essential to scientific enterprise. Burge, however, is sceptical (1986, 9):

. . . I shall assume that individualism is *prima facie* wrong about psychology, including cognitive psychology . . . the generalizations with counterfactual force that appear in psychological theories, given their standard interpretations, are not all individualistic.

Burge summarizes two arguments against non-individualistic accounts. These arguments, according to Burge, presuppose the familiar Twin Earth cases. The first, according to Burge, takes it that twins are *behaviorally* identical, so there is no reason to suppose that we cannot formulate explanatory generalizations to cover them, in spite of the contextual differences that apply. The second argument is essentially the causation argument with which we are already familiar. According to Burge (p.13), proponents of this argument take it that

the determinants of behavior supervene on states of the brain. So if propositional attitudes are to be treated as among the determinants of behavior, they must be taken to supervene on brain states. The alternative is to take propositional attitudes as *behaviorally irrelevant*.

Burge attacks the presuppositions of the first argument, arguing that the conception of behavior that is at work there cannot sustain an intelligible position. We have already made some points with regard to the subproblem of the individuation of *behavior* and its relevance to explanatory role problems, and must put further comment on this to one side here. Instead we will focus on one of Burge's attacks on the second argument; it is reasonable to suppose that these points can be applied to *both* positions.

The focus of Burge's criticism concerns the individuation/causation relation. The argument for individualism in psychology, according to Burge, proceeds by

way of what he calls (p.15):

bland observations about the etiology of mental events and behavior. It is plausible that events in the external world causally affect the mental events of a subject only by affecting the subject's bodily surfaces; and that nothing (not excluding mental events) causally affects behavior except by affecting (causing or being a causal antecedent of causes of) local states of the subject's body . . . only if mental events (and states) supervene on the individual's body can the causal principles be maintained.

But this, he says, is confused reasoning. According to Burge, there is every reason to suppose that the individuation of explanatory entities in psychology must meet causal criteria, but there is no reason to suppose that it follows from this that standards of individuation are *individualist*; that is, intrinsic, local, proximate: *what it is that causes have to be*. Rather, he claims (p.16-17):

. . . there is no simple argument from the causal principles just enunciated to individualism . . . local causation does not make more plausible local individuation, or individualistic supervenience . . . causation is local. Individuation may presuppose facts about the specific nature of the subject's environment.

How does Burge make local causation and non-local individuation compatible? First, he urges care regarding the use of the notion of "affect," a preliminary to the examples that are meant to carry his counter-argument intuitively. The individualism argument proceeds by way of the claim that commonsense conditions on causality impose

certain constraints on all that is involved in mental causality, both in the world-mind direction as in the mind-behavior direction. That is, the head-independent world can only impinge on mental events via the subject's body; and the only way for anything mental to be connected causally to some behavioral episode is for it to affect the body involved in that episode. Twin Earth cases are thought to pose a problem for content - and consequently suggest that psychology needs to be individualist - because the nature of such cases is to drive a wedge in between mental states and states of the body (recall that twins are intrinsically identical down to their molecular structure; but they do not share the same mental states). One way to put this is to say, as Burge has it (p.16): "events in the environment are alleged to differentially 'affect' a person's mental events and behavior without differentially 'affecting' his or her body." This is bad news for causal explanation, clearly, especially when what is supposed to be explained is the connection between what I thought with what I did: so psychology must be individualist in its taxonomy of mental states.

What Burge claims is that this line of argument equivocates on 'affect,' confusing causation with individuation. Events in the environment 'affect' in the sense of *individuate* a person's mental events and subsequent behavior; this does not mean, however, that such

individuation fails to get implicated in someone's psychology according to the usual constraints; it is not the case, that is, that the body is not also causally 'affected.'

Burge offers a number of examples to illustrate his claim that local causation and remote individuation are not mutually exclusive: predominant among them is the notion of *functional* individuation. The relations involved in the movement of tectonic plates are all physical and proceed along the usual causal lines - but what counts as Continent A has something to do with its spatial relations; so that if the very same collection of molecules were to sink to the bottom of the ocean, or be mysteriously removed to Mars, there is no reason to suppose we would continue to count it as Continent A. Likewise, if the very same molecular configurations that in us perform visual functions were to perform *menstrual* functions in other organisms, then, in those organisms, they wouldn't be eyes. These examples are meant to show that we do not always individuate along causal, intrinsically supervenient lines.

The readily available literature is replete with parries and thrusts on the issue of whether psychology must be individualistic;⁴² I shall limit myself to a few remarks.

Burge, who, like others, takes it that his own versions of the TE cases are almost universally applicable

(see chapter 1), fails to realize that content's failure to play a causal explanatory role is a question of more than its failure to supervene. His remarks about the individualism thesis in psychology all presuppose that the real problem stems from, as we said above, the fact that content's failure to supervene pulls it apart from its causal role, since it interferes with a reliable, generalizable, context-independent correlation of contentful states with intrinsic, physical states.

Now, we have already seen (chapter 1) that it is a misformulation of the main issue to suppose that the intrinsic states thought to be held stable betwixt twins must be physical states; quite apart from this, what Burge doesn't consider is the case of *supervening* content with no causal/explanatory role. Recall that the main thrust of Burge's counter-argument against the individualist is that there are non-individualistic modes of explanation - in, for instance, geology, physiology and parts of biology - where there is an appeal to entities that do not supervene on the physical substructure (1986, 19). The explanation adheres nevertheless to causal principles. But now what of the numerous cases of supervening but extrinsically individuated content? As we know, a causal role problem for such content arises for all that it supervenes. The fact that there are ways of individuating non-individualistically - yet causally - won't be of great help with respect to the

problem of weakly external content: and, as we have already seen, there are a lot more concepts vulnerable to the redundancy problem than there are likely to fail to supervene.

In sum, we can object to Burge's position from a quite general point of view. The question arises why it is that extrinsic individuation is defended if causal properties have to be intrinsic. It is hard to see, given the adherence to the usual causal requirements, what motivates extrinsic individuation, save faith in the TE cases - faith that turns out to be exaggerated, as we have seen. Perhaps there is an independent reason for defending extrinsic individuation in the face of having to make it compatible with intrinsic causal requirements; but if so, it is not forthcoming from Burge.

e. Teleology

Burge tries, in a sense, to have it both ways, with limited, if any, success. There is another, less compromising view, associated with Dretske (1981;1988); Millikan (1984); Papineau (1987) and Stampe (1975;1977). Fodor veered close to once endorsing something like it (1984;1985), and McGinn once found something to recommend it with respect to the issues raised by content's explanatory role problem (1989).

This kind of view effects an adjustment of the

individuation/causation relation so to comprise an idea familiar in modern accounts of evolutionary biology - teleology. I want to avoid detailed comment on the prospect for naturalizing content by way of teleology (for an exhaustive consideration, see Millikan (1984); also Dennett (1987)). Rather, I will limit my remarks to the way that teleological theories of content can be thought to recalibrate the individuation/causation relation and what results for the problem of content's explanatory role.

The claim, ultimately, is that citing the content of a mental state is to give its *function*, and this will explain why an organism does what it does. What does it mean for content to be individuated with respect to function? And how is this meant to circumvent content's explanatory role problem?

The idea is this. Things like organs, traits or processes are to be understood to have *functions*. The *proper* function of an entity is understood normatively: what it *supposed* to or *ought* to do. The key point is that the normativity has a relational or environmental element - organisms evolve characteristics in the fullness of evolutionary time in order to survive; so what a given characteristic is supposed to do is to be understood relative to this supreme need of the organism in which it resides. But, of course, an organism inhabits an environmental context, which supplies both what to exploit

and what to avoid in the survival contest. The functions of an organism's constitutive parts therefore must be understood relative to the context in which the organism either flourishes or perishes. In fact, a rather stronger claim can be made: not only are these entities to be understood relative to the environment, they are defined with respect to the environment. So: a duck's foot is webbed so it can cope with water; cat's eyes reflect light so as to be able to hunt in underilluminated conditions; etc. The function of these entities is hence *extrinsically* defined: it is individuated by reference to things outside the organism itself.

Now the interesting and controversial claim with respect to content is that representational mental states - not just physical states - are themselves capable of being understood as having a proper biological function that is determined relationally. This is perhaps most easily seen in the case of *desire*. The desire that *p* has a function - that of bringing it about that the organism does what it has to do to get *p*. Sometimes getting *p* has crucial survival value, and sometimes not (the unfulfilled desire for food and water has earlier, quicker and nastier effects than does the unfulfilled desire for a date with Peter the aerobics teacher). The mental state is nonetheless determined with respect to something independent of the organism that possesses it. The general picture, according to McGinn

(1989, 148), is this:

Evolution must install mechanisms which perform . . . interlocking functions of sensitivity to need and sensitivity to what in the world will meet the need: desire and perception are the solutions it has come up with, at least in 'higher' organisms. The teleological theory sees in these basic relational functions the deep roots of content. Belief comes into the picture as a way of guiding behavior in the light of perception so as to satisfy desires. Thus the function of the belief that there is water here is to combine with perceptions of water to guide behavior in the satisfaction of the desire (and hence need) for water: the desire can only cause the right goal-satisfying behavior if it controlled by beliefs about the current state of the environment. In this way, the, the functions of mental states are absorbed into their content, thus incorporating the worldly items that the functions themselves concern. Teleology is what originally brings the world into the mind. It spans the divide.

Lots of complicated issues converge at this point: whether and how 'indication' is a part of mental representation; misrepresentation, original intentionality and derived intentionality; how it is that genuine intentionality can emerge from relational proper function (or, for that matter, from causal theories) - how, that is, that function can be thought to be attached to something *propositional*; purposes that go forever unrealized.⁴³ Fortunately, the only one that can concern us here is how an individuating criterion of content according to function is thought to overcome the problem of content's explanatory

role.

The solution is thought to lie in the fact that teleologically-individuated content is understood to have a non-causal explanatory role. What explains why organism *O* did *a* is what his mental state *s* was *for*: and this must be understood as *world*-involving, since organisms are fundamentally environment-interactive instruments. The explanation of *O*'s behavior goes beyond mere mechanisms, as his behavioral episodes are interpreted to be goal-directed. Causal mechanisms are at work, to be sure, but, on this view, they turn out to be explanatorily inadequate, since they are just bedrock, so to speak: necessary elements in the process of producing behavior - hence explaining it - but not sufficient; *true* descriptions but not *explanatory* ones. On this view, it is no surprise that there is a risk to content on a causal criterion of explanation, since a causal mechanism can never in itself reveal what it is for. The basic form of psychological explanation on the teleological account thus denies, in an important sense, that explanations are *causal*, as long as causal explanations are interpreted as descriptions of the workings of the causal machinery involved. McGinn summarizes the point (1989, 152-153):

The explanatory property - namely, a state with a certain content - is just not a 'mechanistic' property, any more than functional properties in general are. Reasons are not causally explanatory *qua* reasons, though reasons

may in fact be causes. The causal mechanisms are in the head (literally), but reasons themselves are not in the head, since their contents are not. But in the same way the property of having the function of pumping blood to the muscles is not in the heart, since it involves a relation between the organ that is the heart and the muscles; the mechanism of that function, however, is right there in the heart. Functional properties of this kind are not causally explanatory either - if this means that they figure in the mechanisms that bring about the effects. We might more perspicuously say: explanation by reasons is not a species of mechanistic explanation, i.e. an account of the causal mechanisms whose operation led to the behavior being explained.

As we have already suggested, this sort of view is by no means uncontroversial; but adjudication of it and the other options for a solution to the problem of content's explanatory role is not part of our project here. Rather, we would only draw attention to the way in which this view - at least in principle - undermines the usual explanatory problem thought to arise when content is extrinsically individuated. It is not the case - so this view suggests - that extrinsic individuation necessarily interferes with the explanatory doings of content with respect to behavior. On the contrary, what could be more explanatorily relevant to an organism's transactions than citing the *function* of the implicated mental state? When we know what that state is for, we can also know why the organism in possession of that state did what it did: the explanation, roughly speaking, just falls right out of the function. Functional taxonomy

is, of course, firmly and genuinely extrinsic, and, by denying that causal properties (though applicable) tell the relevant explanatory story, the teleological determination of content assures the compatibility of externalist individuation and explanatory role. The individuation/causation relation is thus, on this view, reconsidered in favor of function rather than of mechanism; with the nature of explanation itself following suit.

f. Rationalization; or the True Nature of Psychology

One line of resistance to the idea that content has no explanatory role to play in psychology, as we saw above, is to insist that there are ways of classifying explanatory entities according to criteria other than the tracing of their causal properties. This line of argument calls attention to the fact that sciences quite often are not solely concerned with identifying the causes of things, and taxonomize accordingly along non-causal lines.

There is, however, another line of thought that centers on the understanding of what we might call the true nature of psychology. This view has it that it is quite wrong to think that psychology is or ought to be interested in causation. Belief-desire psychology, in particular, is not a form of causal theory at all, rather, it is a radically different kind of understanding which involves a notion of *rationalization*. Explanation in psychology isn't

given by causes, but, for instance, by what the reasons were for an agent's doings: why it seemed reasonable to him to do what he did, given his beliefs and desires. And this is to be understood entirely non-causally.

Such a view tends to go some way to undermine the worries about the explanatory role of extrinsic content. After all, as we have seen, these worries devolve on the inadmissibility of content's playing a causally explanatory role, given its remoteness from head-dependent causal reticulations. Extrinsic content, as we have seen, is quite literally not in a position to bear causal-explanatory burdens. Should we deny, however, that there is such a causal role for it to play, then the worries consequently evaporate. None of the usual considerations could count against wide (or wide*) content, since, on this view, psychology is not in the business of giving causal explanations.

Just to establish the atmosphere, we may take a brief look at the issues that appear to be at work in Wittgenstein's thought about psychology.

Wittgenstein's conception of psychology is being fairly astringent. According to Budd (1989, ix-x):

Wittgenstein claimed that the science of psychology is barren and confusion is endemic in it. This is not due, he maintained, to the fact that psychology is a young science that is still struggling to find appropriate ways to investigate its subject matter. The kind of confusion that reigns in psychology

is, he believed, conceptual confusion: psychologists are prone to unclarity about everyday psychological concepts and the sophisticated experimental methods they employ fail to deal satisfactorily with the problems addressed, which are really of a philosophical nature.

Those familiar with Wittgenstein's approach to conceptual problems will recognize his trademark: what is needed is *therapy* or *remedy*; with enough care, conceptual difficulties will be recognized for the misconceptions that they are. Wittgenstein is, of course, not alone among philosophers in maintaining that lots of what goes on in psychology is a waste of time.⁴⁴

Where the contemporary philosopher of mind might, however, want to part company with Wittgenstein is over his conception of mental causation - such as it is. Wittgenstein appears to find nothing troubling about the view that psychological differences need not be based in physical differences; and he seems to deny the idea that mental events (like sensations) cause behavior. Budd notes (1989, 72):

Now it is hard to understand how the inherent suitability of sensations to play a causal role in the production of behavior could be accommodated by Wittgenstein in any other way than by regarding them as being physical events in people's bodies . . . But whatever the merits of the suggestion, it is clear, I believe, that Wittgenstein would not have accepted it.

This is borne out in Wittgenstein's clearly sceptical

remarks concerning the possibility of there being a causal connection between the physiological and the psychological.⁴⁵ Given this, the possibility of explaining human behavior by reference to what happens to people's bodies really seems a non-starter (1981, #608;#609;#610;#613):

No supposition seems to me more natural than that there is no process in the brain correlated with associating or with thinking; so that it would be impossible to read off thought-processes from brain-processes. I mean this: if I talk or write, there is, I assume, a system of impulses going out from my brain and correlated with my spoken or written thoughts. But why should the system continue further in the direction of the centre? Why should this order not proceed, so to speak, out of chaos? The case would be like the following - certain kinds of plants multiply by seed, so that a seed always produces a plant of the same kind as that from which it was produced - but *nothing* in the seed corresponds to the plant which comes from it; so that it is impossible to infer the properties or structure of the plant from those of the seed that it comes out of - this can only be done from the *history* of the seed. So an organism might come into being even out of something quite amorphous, as it were causelessly; and there is no reason why this should not really hold for our thoughts, and hence for our talking and writing.

It is thus perfectly possible that certain psychological phenomena cannot be investigated physiologically, because physiologically nothing corresponds to them.

The prejudice in favour of psychophysical parallelism is a fruit of primitive interpretation of our concepts. For if one allows a causality

between psychological phenomena which is not mediated physiologically, one thinks one is making a profession that there exists a soul *side by side* with the body, a ghostly soul-nature.

Why should there not be a natural law connecting a starting and a finishing state of a system, but not covering the intermediary state? (Only one must not think of *causal efficacy*.)

In short, the kind of conception of psychology arguably exemplified in Wittgenstein is one in which a causal role for mental states is simply not on the cards, at least via mediation by physical states. Whatever the independent plausibility or validity of this view, what concerns us is merely this. The extrinsic individuation of content, as we have taken pains to examine, has problematic *causal* consequences for content's explanatory role. Deny that psychology is a causal enterprise, however, and the motivation for finding an alternative to the extrinsic individuation of content for the sake of folk psychology is undercut.

One thing we might pause to consider is that there is, in fact, a plausible analysis of 'because' that might be adapted to the idea that psychology is not a causal undertaking. Suppose you want to know why a proposition *p* is true. Someone says, "it's because it follows from propositions *q* and *r* from a rule of inference *I*." Now, this use of 'because' is hardly causal; rather, it's *logical*. Certainly nothing in any of this shows that the propositions

involved couldn't have wide or wide* content. So there is a precedent for a non-causal use of 'because,' and the 'because' of psychological explanation might best be thought of as analogous. Maybe the most profitable emphasis concerning the nature of psychological explanation could center on the *logical* relations that hold between mental states and action, with the resulting explanatory role of mental descriptions being that of justifying these relations.

Now there is a conception of psychological explanation, associated with Davidson, that might be thought to have something in common with the more radical Wittgensteinian position, and which touches on the above points. Although Davidson's position on the issue of causal explanation is familiar, what is worth exploring are the implications for such a view on the issues raised by externalism and explanation. So let us briefly explore the convergence of these lines of thought.

The Wittgensteinian position might be thought to be an expression of the idea that there is some incompatibility between something's having a normative or rationalizing role at the same time as its being causal, a view, that, until Davidson's work emerged on the scene, was not without its defenders (see Evnine 1991, for discussion; also Davidson 1980). As we have seen, it is hardly the case that explanation in mathematics, logic, or even fiction is not

genuine explanation; but then it appears to have to be conceded that explanation isn't always causal. Now the objection that can be made to this is that the relation between events - say, a mental event and a physical event - is of such a nature as to impose unavoidable causal considerations; but we know that while causal relations may indeed hold of events, it does not follow from this that an event causally *explains* another to which it bears a causal relation.

Davidson, as is well-known, claims (1980, 3-19) that rationalization is indeed a form of causal explanation (p.3):

. . . What is the relation between a reason and an action when the reason explains the action by giving the agent's reason for doing what he did? We may call such explanations *rationalizations*, and say that the reason *rationalizes* the action.

In this paper I want to defend the ancient - and commonsense - position that rationalization is a species of causal explanation. The defense no doubt requires some redeployment, but it does not seem necessary to abandon the position, as has been urged by many recent writers.

Now, given that there is a certain amount of controversy concerning exactly what, given his arguments, Davidson is *entitled* to claim - quite apart from what it is that he *does* claim - it is best to get the basics right. Rationalization is giving the reasons for why somebody did what she did - the beliefs and desires she had and for which

she did what she did. The question is: is giving reasons a form of causal explanation - is it a form of causal talk at all?

Now - as we see in the quote from Davidson above - Davidson's thesis is, apparently, that rationalization is causal explanation. What complicates the situation is that in his (1980) he argues primarily for the view that reasons cause actions. And the trouble is, the latter does not entail the former.

The entailment doesn't hold for the familiar reasons that 'causation' is referentially transparent, but 'explanation' is referentially opaque. So, as is familiar, it's possible for reasons to cause actions, but not be explanatory under those descriptions: either at all, or more moderately, explanatory under those descriptions but not causally explanatory under those descriptions. There is certainly logical room for the idea that reasons cause actions, and that the property of being a reason, though explanatory, is not causally explanatory, but, instead, has a *sui generis* justifying explanatory role.

Now, Davidson tends to conflate the two points in question (see Evnine 1991 for discussion); arguing that reasons not only cause actions but are causally explanatory of them. Certainly in (1980, 3-19) he takes it that if reasons were causes, then it would follow that rationalization is a species of causal explanation. And what

further complicates matters is the appearance in Davidson's later work of the notion of anomalous monism.

One of Davidson's fundamental claims is that the mental is *anomalous*. Anomalism of the mental is the view that states and events - beliefs, desires, perceptions, sensations - are not covered by either psychophysical or psychological laws; laws that purport to govern, respectively, the relations between mental states and events and physical states and events; and the relations between mental states and events to one another.⁴⁶

Davidson is unfazed by the seeming incompatibility. He notes (1980, 207):

Mental events such as perceivings, remembering, decisions, and actions resist capture in the nomological net of physical theory. How can this fact be reconciled with the causal role of mental events in the physical world?
... I start from the assumption that both the causal dependence, and the anomalousness, of mental events are undeniable facts. My aim is therefore to explain, in the face of apparent difficulties, how this can be.

The strategy Davidson employs is to dissolve the supposed inconsistency of three principles: 1) that mental events causally interact with physical events; 2) that *causality* entails *laws*; and 3) that (p.208) "there are no strict deterministic laws on the basis of which mental events can be predicated and explained (the Anomalism of the Mental)."

The anomalism of the mental results from a number of

things taken together. Briefly, mental events and physical events do causally interact, and such a causal relation can be captured by a true singular causal statement. Causal relations are governed by causal laws. There are no strict causal laws that hold between mental events (*qua* mental) and physical events; or between mental events and other mental events. So mental events need be thought of as having physical *descriptions* under which they fall under causal *physical* laws. But the mental - *qua* mental - is anomalous. And this purports to make it possible for someone to be a materialist without having to violate reasonable intuitions concerning the distinction between mental and physical properties.

Davidson has been criticized for being rather unclear about the combination of the thesis of anomalous monism with his claims that rationalization is a species of causal explanation. In fact, many take it that Davidson is compelled to have to give up the idea that rationalization is a form of causal explanation (Evnine 1992) - in spite of the fact that he himself is explicitly in the business of defending just that connection.

We need to build up to this. Davidson (1980, 3-19), in arguing against the traditional picture that denies that rationalization is a species of causal explanation, argues also against the view that tries to distinguish causal explanation from rationalization on grounds that laws are

involved essentially in ordinary causal explanations, but not in rationalizations. Davidson points out that Hume's doctrine that singular causal statements imply generalizations is ambiguous, and that the weaker interpretation fits both causal explanation and rationalization. And given that Davidson quite explicitly is taking himself to be giving a defense of rationalization as a species of causal explanation, it is clear that he finds in this line of argument support for the view that rationalization is causal explanation.

But this means that Davidson thinks that causal explanation involves laws. And putting this together with his thesis of anomalous monism generates headaches. The familiar criticism runs roughly as follows: if explanation always involves laws, but (by anomalous monism) there are no laws of psychology (of actions), then the psychological explanations got by rationalizations cannot involve laws, so rationalizing explanations are not causal explanations.

Such is a sketch of the background to the Davidsonian position. The thing to consider now is the connection between rationalization and causation in the explanation of behavior.

A behavioral episode is particular event, and only under certain descriptions can such an episode count as an action. Such descriptions involve an *intention*; a *reason* for the production and performance of the event in question.

Standardly, actions are events intentional under a description under which they are *rationalized*; and it is the content of the mental states involved that give the reasons for them. The question I want to focus on is how reasons are related to the actions for which they are reasons. Davidson (1980, 3-19) claims that apart from the relations between content and the description of an action that make the former a reason for the latter, a further, *causal* relation need obtain so as to make a mental state the reason for an action. So, when one says "I refused him *because* he was ungentleman-like" the 'because' is meant, according to Davidson, to be understood causally. Reasons are therefore causes, and reason explanations are causal explanations.

As we have seen, this appears, on the face of it, incompatible with Davidson's claims about the anomalism of the mental - but, while Davidson denies that there are *strict* psychophysical and psychological laws, he does not deny that there exist psychophysical and psychological *generalizations*. So a causal role for content (reasons) in explanation is defended as a viable possibility.

Something in all in of this is worth taking a closer look at, however. We know that not all descriptions of events will be relevant to the explanations of those events. Not all descriptions of actions, thus, will be relevant to the explanations of those actions. Davidson's position appears to overlook a certain possibility. What if the

intentional description of an action isn't causally explanatory?

That this might be so can be seen, preliminarily, in the following way. Davidson's position is that there is a causal relation (1980, 17): "what other relation could there be other than causation?") between content and actions and that citing the implicated mental state is to cite the cause in question. But the question that arises is whether from the fact that events have the property of being the (say) belief that *p*, that it follows that they have a or the causally relevant property. Certainly - and as we have mentioned in our remarks above - it could easily be the case that when you frame an explanation of an event you cite something that is a cause; but it would not follow that you cite that cause under a description that is a causally or explanatorily relevant description.

The upshot of these remarks is that, in spite of Davidson's assertions - and arguments - to the contrary, rationalizing explanation need by no means be or be thought to be causal explanation. A logically possible position is one that claims that although mental states are causes and enter into causal relations, it is not the case that mental descriptions need purport to be causal descriptions - they might instead purport to be rationalizing descriptions, functioning rather in the same way as a description would in the context of a mathematical or logical explanation. So,

for instance, when we describe mental states in explanatory contexts, it is certainly possible that we just might not be interested in them as causes - though they happen to be causes.

He who holds this position is in a position to agree with mental causation, but would not have to agree that mental explanation is causal. He could claim that it is true that if one has a desire and one does something as a result of it, then that desire caused one to do it - it caused one to do it via the fact that something in one's brain token-identical with that desire caused one's body to move. But it needn't follow that the explanation of such an intentional action - given by saying "I did it because I desired to do it" is a *causal* explanation. Certainly there's no entailment between 'A caused B' and 'B because A.' The 'because' might just as easily be interpreted *non-causally*.

This kind of view has - at least in principle - very different consequences for the bearing of externalist considerations on the future of folk psychology. Quite apart from the independent merits or weaknessness of such a view, it is clear that its proponents are under no obligation to cave in to externalist pressures on content-based psychology. On this position one could claim that all causally relevant properties in cases of mental causation are entirely internal to the agent; mental states themselves, on the other hand, are completely external (that

is, individuated relationally). Mental descriptions are indeed explanatory - just not causally explanatory - so this separation needn't be troublesome to content-based psychology, as long as we think of its explanations as non-causal.

The Wittgensteinian position is a radical one: no mental causation. The above position - as yet unnamed - is a more moderate version of the same basic idea. On this view we allow for mental causation, but given that we understand the difference between causal relations and causal explanations, we can concede that there is quite enough logical room to deny that psychological explanation is causal explanation. If it is non-causal rationalizing explanation, then the pressure in the way of doing anything to eliminate content is decreased.

10. Conclusion

In this chapter I have examined two related issues. The formulation of externalism defended in chapter 1 requires a reconsideration of its consequences for content-based psychology. We have seen that the distinction between weak and strong externalism imposes a reevaluation of the nature externalism's consequences for explanation. I have argued that the relational or extrinsic individuation of content poses a general problem for the supposition that content plays a causal role in psychological explanation,

quite independent of content's supervenience on internal states or its failure to so supervene. I have urged a more critical look at the real nature of the failure of supervenience problem, claiming that it is but an instance of the more widespread problem of content's extrinsicness, and hence redundancy, in causally explanatory contexts. I have distinguished the redundancy problem from other - superficially similar - cases of redundancy, and argued that objections to these fail to make the case against the notion of causal redundancy that I argue is the basis of the problem of content's explanatory role.

There are a number of responses to the initial argument that extrinsic individuation and causal role do not a compatible combination make. These I summarize and consider in turn, indicating which have more plausibility than the rest. I have avoided any outright partisanship, preferring to consider the options from a more abstract point of view. It turns out that the presuppositions at work concerning the individuation/causation relation have the result not only of distinguishing the alternatives for a solution to the explanatory role problem in interesting ways, but of shedding some light on the nature of content-based psychology as an explanatory enterprise.

That externalism has consequences for psychological explanation is an issue that is, at least in some sense, been well-worked in the literature. In fact, one could

easily come to the view, given recent and familiar emphases, that externalism has consequences exclusively for explanation and for the issues of self-knowledge and, more generally, scepticism. What I argue in chapter 3, however, is that externalism has further consequences. As we shall see, both the nature of analytic truth and the notion of 'semantic innocence' are subject to redeployment from the extrinsic individuation of content.

NOTES TO CHAPTER 2

1. The *locus classicus* for enquiries of this kind is Hume (1978).

2. Useful but flawed. See Kitcher and Salmon (1989).

3. Laws involving dispositions are notoriously controversial. See W. Salmon (1992).

4. There is a further issue, concerning the distinction between strict and *ceteris paribus* laws, which has received some attention in the literature (see, for starters, the debate between Fodor and Schiffer, 1991). I will be making some remarks about the status of special sciences below, an issue that classically involves that of hedged laws, but I do not anticipate becoming enmeshed in this controversy.

5. Wittgenstein (among others) seems to hold that mental states can't cause anything. See below, section 11.

6. A few clarificatory remarks about behavior and its characterization are indicated: for an excellent introduction to the issues, see Dretske (1988); and for further details, Burge (1986, 10-15 is useful). What explains behavior is relevant to the way it is described. Say, for example, that one presupposes or defends a non-intentional account; an account that makes use of ordinary physical descriptions. Given that non-intentional explanandum, the odds are there will be no pressing motivation to offer an intentional explanans, and vice versa. Dretske puts it like this (1988, p.83):

. . . what we are trying to explain when we advert to such content-bearing entities as beliefs and desires, is not the physical movements or changes that are the normal product of behavior. What we are trying to explain, causally or

otherwise, is not why our limbs move but why we move them.

Now, describing an event as a behavioral event is a (weak) attempt at neutrality. This is because to describe an event as an action presupposes intentionality, something that we might want to avoid. Consider that one issue that could conceivably arise with respect to externalism is this: perhaps a non-intentional redescription of the behavioral event in question might succeed in removing the necessitating conditions for the problem of content in explanation. If the behavior in question is described in ways that are not externalist then it may not need explanation under externalist descriptions of mental states, which have such untoward explanatory consequences. If you describe something as a *reaching for a* then it is likely that you shall have to explain it in terms of having beliefs about *a*. Describing the behavior as a mere *reaching*, however, may avoid the problem entirely.

7. See also Stich (1978, 575-576).

8. The more general the problem, of course, the more, ultimately, we would have to consider abandoning either externalism or propositional attitude psychology. Any views I have about which one should go I reserve for another paper.

9. The *locus classicus* concerning narrow content is Fodor (1987), and most anything of Fodor's from 1987-1992. This notion has generated an industry: see the bibliography to Loewer and Rey (1991).

10. A claim at odds, of course, with the tenor of much of the literature on the subject, which takes failure of supervenience to constitute a general problem for the role of semantic content in psychological explanation. This is just a error, parallel to the original one concerning the head-independence of meaning. Just as Putnam and many of his commentators uncritically took his points to apply to all content, so further work concerning content in explanation perpetuated and compounded the original error by supposing that all content fails to supervene, so that content-based psychology in general was in hot water.

11. Note that adhering to a TE formulation of externalism will interfere with a grasp of this point. As we saw in chapter 1, 'non-supervenience' is but one way of formulating the issue. Many commentators avail themselves of the singular thought formulation: see chapter 1, section 9.

12. McGinn argues (1989, chapter 2) that an interesting parallel argument, in effect, has been given by Field (1980). His argument that science ought to be done without numbers can be understood along these lines: physics is certainly replete with numerical references, but it is scarcely credible that it is *numbers* that are driving the causal mechanisms that are at work in, say, what happens to water when it reaches 0° Celsius. Numbers are abstract, so they have no causal powers, and, even if they *did*, they are *extrinsic* to the entities whose doings they are employed to characterize.

13. See Marcus (1971) for an interesting discussion. It so happens that in Dretske's discussion of the individuation of behavior (1988, chapter 1), he makes use of a distinction between internal and external not entirely devoid of quasi-essentialist points to illustrate his points about systems, their behavior, and action (p.2):

A bee's stinging a child qualifies as bee behavior, as something the bee does, not simply because *M* (penetration of the child's finger by the bee's stinger) occurs. For this can happen without the bee's doing anything - if, for example, the child accidentally pokes its finger with the stinger of a dead bee. This would be a case of some external (to the bee) event's causing *M*. To get bee behavior, to have something the bee does, the cause of *M* (stinger penetration) must come from within the bee . . . the difference between Clyde's *losing* his job (something that happens to him) and his *quitting* his job (something he does) resides in the locus - *in Clyde or in his employer* - of the cause of termination . . . Despite its apparent crudity, the simple contrast between internally and externally produced movement capture the basic idea underlying our classification of behavior. *If* we have a well-defined

ordinary notion of behavior . . . it is,
with a few refinements, equivalent to
internally produced movement or change.

14. An analogous point concerns the role of the truth-conditions of a belief in explanation. It has seemed to some that the truth of a belief is not part of its causal potential, and we can understand this to mean - with no violence to those views - that the truth of a belief is extrinsic, not part of, the relevant causal machinery.

15. In chapter 1 we noted an Anglo/American divide in the literature concerning formulations of externalism; predictably, the same divide reoccurs with respect to the consequences of externalism for psychological explanation. The contenders to take note of on this subject on the British side subdivide as follows: Boer, Evans, McDowell, Peacocke and McCulloch defend the view that, in their terms, folk psychology (at least as it involves singular thought) is Russellian; Carruthers, Noonan, and Segal oppose it.

16. His example makes use of singular rather than general content: it involves the use of a demonstrative. The kind of example in question, however, can be extended - counterfactually - to the case of general content. It is this extension that results in the widespread problem of content's explanatory role.

17. William of Ockham (c. 1285-1349) urged what is known as the principle of parsimony or economy as a methodological principle in explanation: "what can be done with fewer . . . is done in vain with more." Following this advice is meant to result in the elimination of pseudo-explanatory entities.

18. And also, hopefully, to put an end to any further insistence on the point that formulation of the problem of content's explanatory role somehow essentially involves Twin Earth or failures of supervenience.

19. See below for a close examination of what this comes to.

20. See Strawson (1974) and Evans (1982) for an extended treatment of this.

21. Of course there is controversy concerning the interpretation of counterfactuals with impossible antecedents - which doesn't mean there aren't optimistic views. See Goodman (1983) and Jackson (1991), for starters. In any case, consider that if God exists, he exists necessarily; nevertheless it is no violation of the laws of logic or anything else to ponder the question "if God didn't exist - though of course He does necessarily, then . . ."

22. Leaving no stone unturned, consider the still more extreme counterexample: Even if Descartes were wrong and you didn't know that you existed, but you still believed it, this would on the fact of it make no difference whatever to your behavior. I am grateful to Colin McGinn for this point.

23. Obviously, there are puzzles that arise with respect to necessarily existing entities like the empty set - these aren't relevant here. The fact that there is one necessarily existing set doesn't mean that any claim to the effect that if x is abstract then x exists necessarily is true. The only point I'm making here is that there is no legitimate inference from the fact that x is abstract to the fact that x is a necessary existent.

24. And surely it is unlikely that we want to rule out an explanation like "It has puppies rather than kittens because it's a dog," which is both perfectly good and adverts to properties.

25. I believe it is necessary to point out that the abstractness of abstract universals (as an objection to their role as properly externalistic entities) is really a red herring: and it's one, I suspect, that an exaggerated fidelity to TE formulations of externalism tends to support. Any abstract objects you can have beliefs about aren't anywhere (take numbers), but it just doesn't follow that issues to do with externalism can't be formulated with respect to them. As an exercise, compare what Frege would say about mathematical expressions with what a direct reference theorist would say.

26. The problem of generalizing between me and my twin, however, is one that occurs only when she and I do not share the putatively explanatory content-bearing state: but as we have seen, this is a problem limited to two kinds of mental state, and loses much of its threatening aspect - as well as most of its allure - accordingly.

27. This line of thought might be understood to fall under a more general view, which goes as follows: on any understanding of truth-conditions, an explanatory role problem for content is generated; hence explanation must be formulated with reference to the syntactic properties of sentences or internal representations, properties that exist in tandem with semantic ones. No assignment of truth-conditions to propositional attitudes is required to account for their explanatory role.

This kind of view, of course, stands as an objection to an externalist account of truth-conditions, just one of a host of views concerning the nature of truth-conditions (consider: verification-conditions, truth-conditions as sets of possible worlds, coherence-conditions, pragmatic-success conditions). So something needs to be said about the relation between externalism - in particular, weak externalism - and truth conditions, in the light of this kind of view. In particular, we have to consider the question whether the claim that propositional attitudes are externalistically individuated is *equivalent* to the claim that they have truth-conditions. If this were so, there would be nothing particularly unique or interesting about the consequences of *externalism* for content in explanation - it is the wider notion of *truth-conditions* that forms the basis of the problem, on this view.

A few brief remarks will have to suffice. First: it is clear that an assignment of truth-conditions does not entail externalism, since such an assignment need not be made in terms of worldly entities at all (consider an assignment of truth-conditions in terms of images). Second: it is plausible to suppose that there are explanatory role problems for content generated uniquely by externalism, weak and strong; such problems would be specific to externalism and will not generalize to the problems raised by truth-conditions for explanation (although there may be points of contact). This line of thought is explored in this chapter.

28. A more streamlined way of putting the same point: even if the object or property didn't exist, the same behavior might be manifested, so the object of a mental state is irrelevant to producing the behavior. The external object is not part of the causal machinery in operation.

29. The point to focus on here is merely that from the fact that the claims at hand make use of the terms 'counterfactuality' and 'redundancy,' it shouldn't be supposed that what's at issue is motivated in the same way, or amounts to exactly the same thing.

30. See Fodor (1975).

31. I do not wish to be taken to be defending syntax over semantics in explanation here: in fact, I intend to shy away from any adjudication of this and related issues. Here I intend only to sketch the broad outlines of the various orthodox positions concerning content's explanatory role in order to situate the problem as I have formulated it more generally.

32. Lots of interesting issues proprietary to the philosophy of science arise here: the putative distinction between observational and theoretical terms and the notion of a bridge law, to name but two. See Churchland (1986); Hempel (1965); Kitcher and Salmon (1989); Nagel (1961); and Putnam (1975f), for an introduction.

33. Fodor gives the general idea (1988, 9):

Even if psychology were dispensable *in principle*, that would be no argument for dispensing with it. (Perhaps geology is dispensable in principle; every river is a physical object, after all. Would that be a reason for supposing that rivers aren't a natural kind? Or that 'meandering rivers erode outside their banks is untrue?')

Also useful is Salmon, M. (1989) for an overview of the notion of explanation in the social sciences, a discussion that proceeds by way of noting the difficulties with a reductionist program. More specifically, the debate between those who favor autonomy in psychology and those who do not is well-represented in the literature: see, among others, Fodor (1968 and 1975); Putnam (1975), Pylyshyn (1984), and Churchland (1986), respectively.

34. It is particularly important, in fact, to distinguish TRP from both of these, since it is not immediately clear that these two views are so easy to distinguish from each other. On the face of it, it certainly does seem as if defenders of syntax are arguing in at least the same terms as those who object to special sciences. The syntacticians say: you've got two counterpart possible explanations: one in terms of syntax and one in terms of content. You don't lose any generalizations if you move to the 'lower' level (in fact, as Stich has it, you *gain* some), so the 'higher' level is explanatory excess to be jettisoned. That these views can be distinguished is not an issue I will to investigate here.

35. I leave it to the reader, as an exercise, to match the author with the view.

36. There are no doubt a host of versions, refinements, offshoots and tributaries associated with these views, notice of which we shall have to forgo. For the record, new directions have recently emerged: see Fodor (forthcoming, 1994); McGinn (1991).

37. Interestingly, Stich appears to consider the possibility that one way to save belief/desire psychology is to abandon the causal restraint on explanation in which those attitudes are meant to figure. See (1978, 582).

38. Stich has all kinds of reasons for thinking STM is better than RTM at explaining what needs to be explained. See McGinn (1989, 127-129) for a criticism of the view as a whole.

39. Fodor has recently come about and struck out into different waters; see Fodor (forthcoming, 1994), where narrow content is, in effect, repudiated.

40. See Loewer and Rey (1991), for no end of criticism, also Schiffer (in Villanueva, 1992). But see also Fodor, (forthcoming, 1994) for the next, post-narrow content stage.

41. A few comments regarding so-called 'two-factor,' 'dual componency,' 'two-tiered,' or 'dual aspect' theories of content are in order here. Some have argued that it is in

principle implausible to propose a notion of hybrid or divided content; this, I take it, is supposed to hold quite independent of what notions go into the compilation of the hybrid. I myself don't see that there is any in principle objection to a hybrid notion of content, but that is a topic for another project. For a glance at those who agree, see, among others, Field (1981); Loar (1981); McGinn (1982). For those against, see, among others, Lepore and Loewer (1987).

In fact, I am sceptical that any viable solution to the explanatory role problem for content will be possible save in the light of an acknowledgment that the semantic properties of content are extrinsically individuated, but that its causal properties will have to be accounted for in other terms. Fodor, for one, explicitly claims that externalism is most likely right for semantics; what he resists is that this has any untoward repercussions for the role of content - the causal role of content - in explanation. He notes (1991, 6):

Externalism is independent of individualism because, whatever the explanatory status of broad content, it is not in dispute . . . that the content of my twin's water thoughts differs from the content of mine; or that 'water' means something different in my mouth and in his; or that these semantical differences derive from differences in our respective head/world relations.

Many of the proposed solutions to the explanatory role problem make use of a hybrid notion of content, even if implicitly, as we can see in the survey of the solutions in this chapter. What is interesting is that in spite of many protestations to the contrary, many of those who insist that content can't be hybrid, and who insist that criteria of individuation and causation stand or fall together (either internalistically or externalistically) end up, it seems to me, defending just such a hybrid notion of content. On Fodor's own view, for instance, content is most naturally and plausibly thought to be a hybrid: content is composed of narrow and wide aspects, factors, tiers, components. McDowell, for another, is explicitly dismissive of the idea of a hybrid notion of content (1986, 3):

. . . meaning what one does by a natural-kind word has been revealed to be a composite . . . Something's being the state of mind it is cannot be purely a matter of how things are 'in the head,' if meaning what one does by a natural-kind word is an example . . .

[this] protects a form of the conception that has come under pressure: it incorporates a narrow psychological fact in a composite picture of the problem cases. Some version of this is often represented as the only possible way of accommodating Putnam's thesis. But there is, as we have said, a choice . . . Need there be a psychological state of that kind even partly constituting one's meaning what one does by a natural-kind word?

What McDowell calls for is the notion of object-dependent thought with which we were occupied in chapter 1, as the basis for a more general understanding of externalism. McDowell thinks anything less than this conception of content "fails to supply a satisfying account of the mind's directedness towards the world" (p.166). The trouble is, he makes the further point (p.3) that "no doubt what is 'in the head' is causally relevant to states of mind." And how this is supposed to square with his mono-content view is anybody's guess.

42. See, for an introduction, Burge (1986); Egan (1991); Fodor (1988); and Segal (1989). Recently Fodor appears to have partly conceded that things other than intrinsic causal mechanisms can be relevant to explanation - see his (1994), forthcoming.

43. For much discussion on these and related issues, see Dennett (1989); Dretske (1988); Fodor (1990); and McGinn (1989), among others. The bibliographies to these are useful, too.

44. See Fodor (1986, chapter 8).

45. Wittgenstein also appears to be somewhat sceptical about there being a connection between physiological states and processes and mental events, because - somewhat alarmingly - he is not altogether sanguine about the existence of physiological processes at all (1980, #1063):

Thinking in terms of physiological processes is extremely dangerous in connexion with the clarification of conceptual problems in psychology. Thinking in physiological hypotheses

deludes us sometimes with false
difficulties, sometimes with false
solutions. The best prophylactic against
this is the thought that I don't know at
all whether the humans I am acquainted
with actually have a nervous system.

Not to mention that it is "imaginable that my skull
should turn out empty when it was operated on" (Wittgenstein
1969, #4).

46. The claim that the mental should be anomalous, is,
I believe, Davidson's attempt to embrace materialism while
rejecting reductionism.

CHAPTER 3

EXTERNALISM AND ITS CONSEQUENCES FOR SEMANTIC INNOCENCE AND ANALYTIC TRUTH

Leibniz . . . was always engaged in trying to construct such a mathematical logic as we have now, or rather such a one as Boole constructed, and he was always failing because of his respect for Aristotle. Whenever he invented a really good system, as he did several times, it always brought out that such moods as Darapti are fallacious . . . but he could not bring himself to believe that it was fallacious, so he began again. That shows you that you should not have too much respect for distinguished men.

Russell, *The Philosophy of Logical Atomism*

1. Introduction

In the philosophy of language (as well as in related metaphysical inquiries) there is a long tradition of distinguishing meaning and reference. Theories of meaning and theories of truth; intension and extension; concept and object; connotation and denotation - not only are these not to be confused, but, depending on one's ideological presuppositions, one is thought to be in better shape theoretically than the other.¹

This canonical ontological distinction is most explicitly captured in Fregean semantics, where it is put to work to solve an apparent puzzle concerning the semantics of

expressions in intensional or so-called opaque contexts. It is also clearly what is behind the common-sense distinction between analytic and synthetic truth: the distinction, as it is often and most simply described, between truth in virtue of meaning alone and truth in virtue of meaning and the world.

Externalism about meaning, as we are now in a position to appreciate, reconfigures the distinction between meaning and reference. No longer is it possible to viably posit a conceptual realm, in some sense co-existing with and related to that of the ordinary world of objects, but autonomous or independent from that world. The world of reference now permeates the realm of meaning.²

We have seen that this account of meaning has decided repercussions for mental content and the explanation of behavior. In this chapter we trace the consequences of externalism for the semantics of expressions in intensional contexts, and for an account of analytic truth. The issues that concern us here are simply - the simplicity is slightly deceptive - put. First: given, as externalism about meaning demands, that the meaning of an expression is (at least in part) its *reference*, a minute but interesting consequence results for a Fregean account of the reference of expressions in opaque contexts. Such an account has been thought to violate 'semantic innocence' (Davidson 1980), in its claim that the reference of expressions in indirect

discourse are the ordinary *senses* of those expressions: the senses they bear when outside of those contexts. The reference of an expression is an ambiguous thing, therefore, on Frege's account.

What I claim here is that one - unnoticed - consequence of externalism is to make it possible to recapture semantic innocence for a Fregean account of opacity. If meaning is determined by reference, then, when an expression refers to its ordinary sense in an opaque context, it - indirectly, perhaps - really preserves its ordinary reference. Semantic externalism makes it possible to deny that meaning, senses, intensions and the like are corrupt and exotic creatures, reference to which is perilous and unjustified by any properly hardheaded and virtuous semantics.

Externalism has further repercussions for the question of the nature of analytic truth. This is an issue of monumental controversy, much of which we can very fortunately set aside. What I am concerned with is this: analytic truth has traditionally been conceived of as purely conceptual truth with certain concomitant modal and epistemic properties - purely conceptual in the sense of independent of the world. As we know, externalism forbids that kind of conceptual autonomy, so the question arises: does externalism at last spell the end of analyticity? I argue the contrary: not only is it possible to give a

perfectly viable account of analytic truth from an externalist perspective; externalism also forces us to reconsider the very nature of analytic truth.

It is worth pointing out that both of these issues lead, often ultra-precipitously, into unsavory, and tangled, philosophical webs. Any number of presuppositions have to be made - and kept in mind - to keep the scope of my remarks from becoming unmanageable. I have indicated these presuppositions, where appropriate.

PART ONE
SEMANTIC INNOCENCE

1. Introduction

Semantic innocence - the expression is due to Davidson (1984, 93-108) - is a view about the reference of expressions. It concerns, roughly speaking, not only the idea that the reference of an expression ought not to be ambiguous, but, in particular, that the reference of an expression should not be thought sometimes to comprise an intensional entity, and, the rest of the time, the mundane and ordinary middle-sized world.

Fregean semantics has been accused, by Davidson, of violating semantic innocence; indeed Fregean semantics is perhaps a paradigm case of such violation. Frege's strategy in accounting for the behavior of expressions in indirect discourse is well known. Frege maintains the principle that the truth-value of a sentence must remain unchanged when one of the expressions in the sentence is replaced with another with which it is co-referential. Expressions in certain now familiar contexts, however, appear to flout this principle. When, for instance, an expression in the scope of a verb of propositional attitude such as 'believe' is replaced with

another whose reference is identical, the truth-value of the sentence seems subject to variation. Thus, although 'Superman' and 'Clark Kent' co-refer, replacement of 'Clark Kent' for 'Superman' in

(1) Lois Lane believes that Superman is a hero

apparently results in a variation of truth-value of that sentence. Frege's solution was to conclude that expressions occurring in opaque contexts do not have their ordinary reference - rather, it is the ordinary sense of the expressions that, in such cases, comes to serve as their reference. Replacing 'Superman' with an expression that expresses the same sense in the sentence ('the defender of good from Krypton,' say), does not appear to have the same effect on the truth-value of a sentence as does replacing the term with one whose reference is the same but which has another sense. Frege thus preserves the principle of substitutivity; no variation in truth-value is held to occur between two belief sentences whose embedded that-clauses express the same sense.³

2. "Two Provinces so Fundamentally Distinct":

Meaning and Reference

Davidson's comment on the Fregean account is notorious (1984, 93-108):

Since Frege, philosophers have become hardened to the idea that content-sentences in talk about propositional attitudes may strangely refer to such entities as intensions, propositions, sentences, utterances, and inscriptions. What is strange is not the entities . . . but the notion that ordinary words for planets, people, tables, and hippopotami in indirect discourse may give up these pedestrian references for the exotica. If we could recover our pre-Fregean semantic innocence, I think it would seem to us plainly incredible that the words 'The earth moves', uttered after the words 'Galileo said that', mean anything different, or refer to anything else, than is their wont when they come in other environments.⁴

Davidson alludes here to a familiar distinction: that between intensional entities and more mundane objects of reference. This distinction may be understood as based on another, one that Quine (1980, 130) has described as:

two provinces so fundamentally distinct as not to deserve a joint appellation at all. They may be called the *theory of meaning* and the *theory of reference*.

The difference between the two theories is revealed in the separate and distinct ontologies to which each permits reference. The subject-matter of the theory of meaning includes meaning, synonymy, and analyticity; that of the theory of reference, instead, includes naming, truth and extension.

Now, Davidson is right: it is difficult to accept that expressions in opaque contexts are ambiguous in the way Frege's account suggests. At the same time, we might suggest

that the temptation to think of such a shift of reference as particularly significant is strongest inasmuch as we think of theories of meaning and reference as discriminated according to tradition. Given that distinction, the reference shift that expressions in a that-clause appear to experience seems to be precisely a shift from entities belonging to the theory of reference to those belonging to the theory of meaning, these being disjoint provinces.

But we no longer tend to think so much of meaning and mental content as - in themselves - harbingers of the exotic. Familiar claims concerning extra-individual environment, both particular and abstract, have taught us that neither meaning nor mental content is isolated from the world of pedestrian reference in quite the simple way Davidson - in keeping with tradition - suggests. And if this is so, there is another alternative to consider in our account of the semantics of expressions in indirect discourse. Semantic externalism offers just the antidote we need to quash the temptation to separate meaning and reference and to provide an alternative account.

3. Externalism and the reference of that-clauses: exotica or the mundane?

We will consider expressions of the form 'x believes that p,' and argue that the determination to reject a Fregean solution to the problem of the reference of

expressions in an opaque context varies in intensity relative to the perceived inextricability of the link between meaning and the extra-individual environment. If the link is perceived as negligible - if a theory of meaning is internalist - then we can claim that the view that expressions in indirect discourse take their ordinary sense as reference will rest on a distinction with some significance *in that theory*. Their non-standard behavior is based, in that case, on a theory of meaning that holds that the reference of a that-clause is determined essentially by sub-cutaneous facts about an individual; on a theory of meaning, that is, that carves a significant ontological distinction between what serves as the sense of an expression, and what its reference. This may well, with some justice, stiffen the resolve of opponents to a Fregean theory, given the deep referential ambiguity that is necessarily visited on expressions on such an account.

If, on the other hand, the link is perceived as essential - if, that is, what determines the meaning of words and the reference of that-clauses is held to be the extra-cranial environment - then we can argue that there will be no in principle rejection of a view that takes the reference of that-clauses to be their ordinary sense; for, on such an externalist view, the semantic values of expressions in a that-clause will be their ordinary references after all, or something that involves them

essentially. There is nothing exotic about an expression's taking either its denotation, or what McDowell (1984, pp. 283-294) has called its *de re* sense as reference in an opaque context, for, on an externalist theory, the cleavage between meaning and reference that Davidson appears to be presupposing does not obtain.

Now, there is another context for which the same point can be made: the context generated by the prefix *It is analytic that*. . . . An examination of this context will show that analytic sentences are another class of sentence whose constituent expressions appear to invite the idea of a shift to their ordinary sense as reference.

Suppose '*p*' is the analytic truth 'All bachelors are unmarried males,' and compare the truth-conditions of '*p*' with those of 'It is analytic that *p*.' What makes 'It is analytic that *p*' true, on a Fregean account, are the senses referred to by '*p*''s constituent expressions, since, in this case, '*p*' is opaquely embedded. Yet what makes '*p*' itself analytically true are the senses expressed by its constituent expressions - for '*p*' is true, as we say, in virtue of its meaning alone. The truth-conditions of '*p*' itself thus appear to involve the very entities relevant to the truth of 'It is analytic that *p*'; the very entities referred to by '*p*' when '*p*' is embedded in a that-clause.

This seems to correlate with the distinction we tend to accept between the nature of analytic truth and that of

synthetic truth. For what is it that distinguishes the truth of 'p', above, from 'q', the synthetic truth 'All bachelors are happy'? The truth-conditions of 'q' are a function of the ordinary reference of the expression 'bachelor' and whether or not the predicate 'happy' applies. But 'p' is held to have a set of unique features⁵ that do not appear accountable for unless 'p''s truth-conditions are specified in terms of the senses of its constituent expressions. These features appear to depend on 'p''s truth-value being determined, as we say, independent of the world; in virtue, that is, of 'p''s meaning, and that alone.

Thus it appears that both the prefix *It is analytic that*. . . . and the prefix *x believes that*. . . . have the same influence on an embedded sentence; both, we may say, are *reference-shifting* operators. This suggests that the distinction we tend to accept between analytic and synthetic truth seems to be yet another distinction supported by the divide between meaning and reference described by Quine above. And, if all this is right, then the same consequences of externalism about meaning that we have claimed for belief will hold for analytic truth. It would seem difficult to deny that whatever is problematic about expressions shifting their reference to exotica in belief contexts could be any less so in the case of analytic contexts. So, although analytic sentences are sentences with distinguishing features of a particular kind, and, on the face of it, it

does not appear possible that an account of their truth-conditions could fail to include an account of their constituent expressions' experiencing a shift of reference to their ordinary senses, nevertheless, we cannot accept reference to exotica for analytic sentences yet deny it for belief sentences. If our objective is to recover semantic innocence, it appears that we must expunge reference to exotica from an account of the truth-conditions of analytic sentences as well.

Let us consider what it is about internalist theories of meaning that encourages an ambiguity-favoring account of the semantics of expressions in the opaque contents generated by *x believes that. . . .* and *It is analytic that. . . .*, and let us see whether externalism about meaning does indeed deflate the "plainly incredible" view that expressions shift their reference to exotica when embedded in opaque contexts.

Belief, we are accustomed to saying, is a relation between an individual and a proposition with constituent structure. Different views assign different structures to propositions, largely as a function of what entities are accepted as relevant to an account of meaning and content. Thus, for example, the semantic analysis of (1), above, will vary, among other things, according to whether the relevant relation is held to obtain between Lois and a proposition whose constituents are the senses expressed by the singular

term 'Superman' and the predicate 'hero'; or whether, instead, Lois is held to be in the belief relation to a proposition whose constituents are the very objects and properties denoted by those expressions.

We are not entirely unaccustomed to discussing analytic truth in these terms, and it is clear that similar considerations will apply to analytic contexts as do to belief contexts. In the context *It is analytic that p*, 'p' also expresses a proposition with constituent structure. An account of this structure will vary in just the same way as do accounts of the constituent structure of a proposition in a belief context.

The contribution made by the reference of expressions is what is crucial to determining the truth-value of a sentence formed by those expressions. Now, an internalist about content must claim that it is internal facts about a subject that essentially determine that subject's mental states. Her belief about Superman's heroism is quite independent of any environmental variations with respect to Superman. The mental state enjoyed in belief is, as we have said, commonly analyzed as a relation to a proposition. And an internalist has to say that the expressions embedded in the *that*-clause that expresses the proposition believed express concepts that have their essential nature fixed in isolation from the environment that contains the references of those terms. In transparent contexts, this split between

what words refer to and what they mean is not as evident as it becomes in opaque contexts. The truth-value of a sentence in a transparent context is a function partly of the meaning of its words and partly of the properties of what those words pick out. But the contribution to truth-value of the ordinary reference of expressions is a liability in opaque contexts; for there, although expressions co-refer, substituting one for the other appears to affect truth-value.

And, at least for the internalist, there is an alternative. Denotation is only one aspect of expressions; they also have meaning, *which is independently determined*. Semantic internalism, operating as it does with a comparatively loose connection between what determines the meaning of expressions and what it is that those expressions refer to, is a theory of meaning where there exists the theoretical option of invoking something other than the ordinary reference of expressions to act as their reference in opaque contexts - namely, their meaning, which has distinct enough properties to offer a solution to a problem generated in those contexts by their denotation.

No such theoretical option exists, however, for an externalist about meaning, whose theory does not carve a decisive individuating incision between meaning and the world beyond the subject. But this implies that to hold that the reference of expressions in opaque contexts is their

sense is, in effect, for the externalist to meet Davidson's plea. Expressions in indirect discourse ultimately take nothing other than their ordinary reference as reference, if the meaning of an expression is essentially determined by what in the environment it refers to.

4. The Varieties of Externalism:

The Account Applied

Within the externalist camp theoretical commitment varies.⁶ Theories of direct reference (TDR), one variety of externalism, hold that the meaning of an expression is identical to its denotation.⁷ Another variety employs, instead, the idea of a *de re* sense. *De re* senses enmesh particulars and properties, so to speak; they are not bearers of them in such a way as to be intrinsically independent of what they bear. What consequences could follow from these types of view in holding that the reference of an expression in an opaque context is its sense - since sense is determined by reference?

To speak of expressions in opaque contexts as referring to their senses is, on a TDR, to speak in a long-winded, perhaps partly ironic way - the fact remains that the reference of expressions in those contexts is their ordinary reference and nothing more.⁸ Further, the meaning of the embedded sentence in (1) does not change, since both 'Superman' and 'Clark Kent' refer to the very same

individual. These are terms with the same sense because sense is exhausted by reference and they are terms with the same reference.⁹

If the meaning of an expression is held instead to consist in the *de re* sense it expresses, its reference in an opaque context will be ultimately no less pedestrian. It is true that, strictly speaking, the reference of an expression in an embedded sentence is, on this view, a *de re* sense, which is not *identical* to its ordinary reference. However - and this is the key point - *de re* senses are essentially determined by and not independent of the objects and properties that are their bearers. If we think of these senses as media through which ordinary reference is preserved, then it is clear that, on such a view, it is the ordinary reference of an expression that genuinely serves as its reference or semantic value in an opaque context. Crucially, a *de re* sense is not to be understood as an entity unmoored from the world of pedestrian reference; and it is precisely such unmoored entities that Davidson repudiates as *exotica*. In these terms, then, we may say that a Fregean account of (1) is innocent of reference to *exotica*.

The apparent variation in truth-value between (1) and (2) is accounted for, standardly, by the claim that expressions can share reference but differ in sense. A particular individual is the reference of the terms in the

embedded sentences in (1) and (2) and it is he who, on this sort of view, determines the sense of those terms. Since a *de re* sense is not identical with the reference of an expression, however, it will not follow that two expressions with the same reference have the same (*de re*) sense. The troubles Lois generates by instantiating (1) and not (2), in spite of the fact that she is held to be in the belief relation to the very same singular proposition, are explained by taking this into account.

5. Opacity, Belief and Analyticity

Now, given the analogy we have pressed between belief contexts and analytic contexts, the same account must hold, with no untoward consequences for analytic truth, for

(3) It is analytic that Hesperus is Hesperus.

But if 'Phosphorus' replaces the second occurrence of 'Hesperus' in (3), (3) seems, *prima facie*, to undergo the familiar variation in truth-value. Suppose, however, that we accept a TDR; if so, we do not seem to be able to deny that it is also true that

(4) It is analytic that Hesperus is Phosphorus.

A TDR holds that meaning is identical to reference; on

this view, *there is no change in meaning* between the embedded sentences in (3) and (4). Further, since meaning is identical to reference, neither (3) nor (4) is any less true in virtue of meaning; although, again, to put it like this is to speak in roundabout way. It is worth mentioning here, that a sentence can be true in virtue of meaning in two ways: in a purely semantic sense, where the analytic truth of a sentence follows from the semantic values of the terms in the sentence, which are assigned to terms by semantic rules alone; and in a more psychological sense, where analytic sentences are held to be those sentences whose terms can be substituted for one another *salva veritate* in a belief context. Accepting (4) as true may grate less, perhaps, if we think of the embedded sentence as analytically true in the purely semantic sense. If instead we characterize analytic truth in the more psychological sense, then facts about the believer become relevant, and the apparent variation in truth-value between (3) and (4) can be explained by taking facts about the believer into account.

If, on the other hand, the embedded expressions are taken to refer to their *de re* senses, then, again, the key point is that the expressions do not genuinely refer to *exotica*, for, again, *de re* senses are not independent of the objects and properties that are their bearers. *De re* senses are, we may say, *object-involving*, if they are expressed by

singular terms, and *property-involving*, if they are expressed by predicates. Thus, in (3) and (4), the embedded expressions refer to their object-involving *de re* senses. This view does not, however, commit us to the truth of (4), since we can allow, standardly, that two terms that share reference need not share the same (*de re*) sense. Further, in

(5) It is analytic that bachelors are unmarried males

we say that the embedded expressions refer to their *property-involving de re* senses: that is, the ordinary reference of these expressions is what individuates the senses they express; via the *de re* senses, thus, the ordinary reference of these expressions is preserved in an opaque context. Further, we can accept the truth of (5) in the usual way: two expressions can share both reference and sense.

6. Conclusion

The context generated by *It is analytic that. . . .* is one for which it may seem as though semantic innocence is the last thing we ought to want to recover. If expressions in this context refer to their denotations or to their *de re* senses, then, given our earlier comparison of the truth-conditions of analytic sentences and those of sentences like (5), the truth-conditions of analytic sentences will be

specified, in the end, in terms of the world - just like synthetic sentences. Externalism about meaning may indeed block the theoretical option of invoking a realm of meaning alone, distinguished by a unique set of properties, to serve as what it is that the truth-conditions for this class of sentence will involve. But, I would insist, it is far from clear that it follows from this that there is no analytic-synthetic distinction.

I have argued that the theoretical alternatives offered by externalism about meaning appear to meet Davidson's plea against semantic corruption in an account of the reference of expressions in indirect discourse. And I have urged that our recuperation of pre-Fregean semantic innocence cannot be half-hearted. What goes for one opaque context must go for all opaque contexts. This suggests that our recovery will be complete only when we overcome the idea that analytic truth is an exotic kind of truth.

PART TWO
ANALYTIC TRUTH

1. Externalism and Analyticity

Quine's scepticism regarding the notion of analytic truth presupposes a - quite traditional - ontological distinction between intensions, meanings, or concepts, and the extra-semantic world. This is manifest in some preliminary remarks of Quine's (1980, 21):¹

Kant's intent . . . can be restated thus: a statement is analytic when it is true by virtue of meanings and independently of fact . . . Meaning, let us remember, is not to be identified with naming. Frege's example of 'Evening Star' and 'Morning Star', and Russell's of 'Scott' and 'the author of *Waverly*' illustrate that terms can name the same thing but differ in meaning . . . It is indeed a commonplace in philosophy to oppose intension (or meaning) to extension, or, in a variant vocabulary, connotation to denotation.

Quine's rejection of the notion of analyticity can be understood as the claim that there is no reason to suppose that a purely conceptual or linguistic element of truth can be made sense of (1980, 36):

It is obvious that truth in general depends on both language and extralinguistic fact . . . thus one is tempted to suppose in general that the

truth of a statement is somehow
analyzable into a linguistic component
and a factual component. Given this
supposition, it next seems reasonable
that in some statements the factual
component should be null; and these are
the analytic statements. But for all its
apriori reasonableness, a boundary
between analytic and synthetic
statements simply has not been drawn
. . . .

Recent developments in the theory of meaning seem to add further support to Quine's rejection of the analytic/synthetic distinction. Semantic externalism claims that neither meaning nor mental content can be determined independently of an individual's environment. Given externalism, to say that a sentence is true in virtue of meaning is to say, ultimately, that it is true in virtue of reference. And if all sentences are true in virtue of the world, it is difficult to see how we can even in principle distinguish some sentences from others as true in virtue of the world-independent, purely semantic entities that their truth-conditions involve.²

I believe, however, that semantic externalism exposes and challenges an uncritical assumption about the nature of analytic truth, one grounded in the traditional distinction between meaning and reference. According to this assumption, analytic truth is some species of linguistic or 'purely conceptual' truth. Quine is of course quite right to reject a distinction between sentences whose truth-conditions involve a linguistic or conceptual component alone, and

those whose truth is settled with respect to the world. But, as I will argue, it is far from clear that it follows from this that there is no analytic/synthetic distinction.

What follows is an account of analytic truth from the perspective of semantic externalism. The truth-conditions of analytic sentences will not be distinguished from those of synthetic sentences; all, sentences, given externalism, are true in virtue of the world. It will emerge, however, that the truth-conditions of an analytic sentence are to be distinguished from what makes it *analytic*. This account is one that will revoke the customary world-independence of analytic sentences - but will not sacrifice their unique modal and epistemic features.³

A programmatic proviso should be noted at the outset. These remarks are not intended as an argument for analyticity; such an argument is a project to be pursued independently. Rather, the notion of analyticity as truth-value in virtue of meaning alone is here presupposed,⁴ in order to reveal an interesting consequence of semantic externalism.

2. Truth-Conditions and Analyticity-Conditions:

A Distinction

As we have argued (see chapter 1), externalist claims can be distinguished as either weak or strong. An account of analytic truth given strong externalism would, obviously, be

unnecessarily restrictive, since, arguably, only a limited number of concepts are capable of strongly external individuation. Accordingly we will develop an account of analytic truth along the lines of weak externalism, and go on to consider some objections.

Two questions must be carefully distinguished: first, by reference to what are the truth-conditions of sentences like

(1) Bachelors are happy

and

(2) Bachelors are unmarried males

to be given? What is it, that is, that determines the truth-value of those sentences? Second, what is it that makes (2) analytic? In what terms, that is, are its analyticity-conditions to be made?

With Frege, we will say that the constituent expressions of a sentence have both sense and reference. We take the reference of singular terms to be individual particulars, and those of predicates to be properties (parting company with Frege). Further, we take it that properties are characterized by what we will call 'structural complexity,' to be defined shortly. Now, it is

objects and properties that are meant to individuate the senses of expressions given externalism about meaning. How does this picture of meaning constrain an account of truth in virtue of meaning alone?

The answer to the first question above need raise no especial difficulties. If meaning is determined by reference, then (1) and (2) have their truth-value fixed by the reference of their constituent expressions.⁵ We can focus on sentences like (2) for brevity. Such sentences are composed of general terms whose reference, as we have said, are properties. Now, given the characterization of a property as a mind-and-language independent entity with structural complexity of varying degree, we can begin to construct an account of the features that contribute to the semantic profile of those sentences we call analytic truths.

3. The Structural Complexity of Abstract Properties

First, we understand the structural complexity of a property to comprise: 1) its identity with other properties; 2) its component structure - the simpler properties of which it is composed; and 3) the relations it bears to other properties. We take a property to have the structural complexity it has necessarily,⁶ and to bear either necessary or contingent relations to other properties in virtue of that structure. Accordingly, we say that the reference of the expression 'bachelor' in (2) - the property

of being a bachelor - has structural complexity. One of the properties with which it is identical is the property denoted by the expression 'unmarried male.' Thus, (2) is true in virtue of the (improper) componential relation between the reference of its constituent expressions.⁷ So much for what makes (2) true - what is it that makes this sentence an analytic truth?

4. Property-revealing sense

In order to answer this question, we need to build up to a notion of *property-revealing sense*. The best way to understand this notion is by way of McDowell's notion of a *de re* sense.⁸ Segal gives an useful and elegant summary of the idea (1989, 41-42):

From Russell comes the idea that certain singular thoughts are object-dependent, in the sense that the object of thought (the object that the thought is about) is itself a feature of the thought's content . . . McDowell incorporates the Russellian idea of object-dependence . . . into a basically Fregean conception of thoughts. A Fregean thought contains only Fregean senses as constituents. A Fregean sense is a mode of presentation of an object of thought, a way in which the object is presented in thought. Since any given object may be presented in different ways, a sense cannot be identical with the object it presents. The point of holding that thoughts are made up of senses is, of course, to explain how it is possible rationally to hold conflicting attitudes towards the same object . . . The synthesis of Russellian and Fregean views is achieved by allowing that Fregean senses may themselves be object-dependent. The

thought that Viola is male contains, in subject position, a mode of presentation of Viola. But that mode of presentation itself owes its existence and identity to Viola. That sense could not have presented a different object, nor could it have failed to present any object at all. The object of thought thus figures in the thought under a particular mode of presentation.

Now, what we are claiming is this: while all senses are individuated by extrinsic entities - properties, according to weak externalism - it is not the case that all senses will be such as to reveal the structural complexity of the properties that individuate them. We will say, accordingly, that while all senses are *property-involving* (that is, *de re*, in McDowell's sense), not all senses are *property-revealing*. Property-revealing senses manifest both the (proper and improper) structural composition of the properties that individuate them and the necessary or contingent relations borne by those properties.⁹ And an analytic sentence, we will say, is one whose constituent terms express senses that reveal the necessary relations borne by the properties denoted by the terms.

How is the property-revealing function of sense to be understood? This is not so easy to say. Nevertheless, a number of clarificatory remarks can be made.

First, the property-revealing character of the sense of an expression is not a direct function of any syntactic complexity in the expression itself. One cannot, that is, distinguish the property-revealing sense of an expression by

simply looking at the syntactic structure of that expression. Any correlation there is a contingent one.

We can enlarge on this point by considering some contrasting cases. First, take a predicate with, as it happens, syntactic *and* semantic structural complexity: 'is an unmarried male.' Now compare the predicate 'is a bachelor.' Here there is no obvious syntactic complexity, but it would be difficult to deny its semantic complexity: it has the same complex sense as the syntactically complex 'is an unmarried male.'

It is not the case, further, that the property-revealing function of a sense is a trivial one. Imagine a case where an individual - perhaps a non-native speaker of a language L - is aware that her community uses a certain word, say, 'bachelor.' She herself begins to use this word, but the descriptive content associated with the word by the other speakers in the community is, for some reason, not available to her. She would thus simply be using 'bachelor' as a label or a name she has picked up from others without knowing any descriptive information about its bearer.

Now, if we extend the claims of a theory of direct reference - one variety of externalism - from singular terms to predicates, it is clear that we may say that for such a speaker, the predicate 'bachelor' is property involving but not property revealing. The predicate, in her idiolect, will have sense; but not *descriptive* sense.¹⁰

Thus, while it is true that it is on the semantic complexity of terms that analytic relations depend, it is clearly the case that not all terms will generate analyticities in the case of particular speakers. It is not sufficient that an expression has as its reference a property with constituent structure which is taken to individuate the sense of that expression in an essential way; for a speaker may fail entirely to associate any descriptive content to that expression.

Nor, further, is it the case that a term expresses a property-revealing sense context-independently. Whether the sense of an expression reveals the structural complexity of the property that, on this view, individuates it, is a matter of its relation to the other senses with which it co-occurs in a proposition. An expression has sense and reference, and such reference has constituent structure, but this will not necessarily be revealed by the sense of that expression on its own (among other things, it would be difficult to see how there could be different senses associated with the same expression if this were so). Instead, we want to say that the structure of the properties that individuate senses may be revealed in the relation of one sense to another. The structural complexity of a property, something that may remain hidden when one considers an expression and its sense and reference in isolation (consider 'water') may better emerge in the

encounter of senses in a proposition.

Thus it is that the relation of senses to one another will sometimes be such as to manifest the structural complexity of the properties that individuate them, by displaying a property's (proper or improper) relations to another in virtue of its structure. It may help to consider that while it is true, following Frege, that sense is the mode of presentation of reference, it does not follow that a mode of presentation will illuminate the structural features of the reference of the expression, by which it is individuated.

It may be useful, in this context, simply to think of the distinction between property-revealing and property-involving sense to be analogous to the distinction between directly referring to an object and referring to it via a description. A descriptive singular term obviously contributes, in the act of reference, much more information about the entity referred to than does a mere label.

So, given an externalist individuation relation, we will say that senses sometimes mirror the structural relations of the properties that individuate them - by bearing those relations themselves. And, if all this is plausible, we are now in a position to entertain an account of analytic truth made in terms of property-revealing sense.

Property-revealing senses account for the modal features that characterize analytic truth as follows. An

analytic sentence is necessarily true in that 1) it is true in virtue of the reference of its terms and 2) the objects of reference of the terms of analytic sentences are, as we have said, properties, which have the constituent structure they have necessarily, and which bear a number of necessary relations to one another as a result. So (2), above, is necessarily true on the view we have sketched, in that what it is to be an unmarried male is part of the necessary structural complexity of what it is to be a bachelor. The structural composition of the property of being a bachelor, we may say, necessarily comprises the property of being an unmarried male: the properties in this case are identical. What may come as some surprise is that analytic sentences will be, on this account, necessarily true *de re*.¹¹

Property-revealing senses further account for the unique epistemic features characteristic of analytic truths. How is it that a sentence composed of terms that express property-revealing senses is a sentence knowable *a priori*? Recall that, given weak externalism, senses are individuated by properties and their relations; all senses are property-involving. For one class of sentence, however, the properties referred to by the constituent terms of the sentence will bear a (*de re*) necessary relation to one another, and, as we have said, a necessary relation will be borne by the senses individuated by those properties themselves, if those senses are property-revealing. No other

appeal will be needed to grasp the structure and the relations characteristic of the properties that determine the truth-value of the sentence. Grasp of the (property-revealing) senses is sufficient. Propositions whose constituents are property-revealing senses are thus both necessarily true and knowable *a priori*. The sentences that express such propositions are analytic sentences.

5. The Account Applied

The following examples serve to illustrate this account of analytic truth:

(1) 'H₂O is partly oxygen.' Here the properties referred to by the terms bear a necessary componential relation to one another: it is part of the structural complexity of the property of being H₂O that it is partly composed of oxygen. The senses of these expressions reveal this structural complexity by bearing a necessary relation themselves. Grasp of the senses is sufficient for knowledge of the truth of the proposition. The sentence is, accordingly, an analytic truth: necessarily true and knowable *a priori*.

(2)¹² 'A fortnight is a period of 14 days.' Here the properties referred to by the terms bear a necessary component relation to one another: the constituent structure of the property of being a fortnight is such that being a period of 14 days is a necessary (improper) part of it. The

senses of these expressions mirror the necessary structural complexity of the properties denoted by the expressions, and grasp of these senses is, again, sufficient for the knowledge of the truth of the proposition expressed by the sentence. The sentence is, accordingly, an analytic truth.

(3) 'Water is partly oxygen.' The senses of the expressions in this case are property-involving, but, while the properties denoted by the expressions bear a necessary relation to one another in virtue of their component structure, the senses fail to capture it; they are not property-revealing. The structural complexity of the properties - crucially, the necessary relations borne by the properties - is not graspable by mere grasp of the senses of the terms in this sentence. Accordingly, we say that the sentence is a necessary but synthetic sentence.

(4) 'There are puddles of water on 73rd street.' In this case, the senses of the expressions can be said to property-revealing, in that they reveal that the constituent structure of the properties referred to by the terms of this sentence is such that they bear only a *contingent* relation to one another. The senses mirror that contingent relation themselves; but grasp of the senses alone is insufficient to know the truth of the proposition. The sentence that expresses the proposition made up of these senses is a synthetic sentence, contingently true.

6. Objections and Replies

We may now turn to some of the objections that might be raised to the foregoing account of analytic truth from an externalist perspective. We will be chiefly concerned with replies to what seem to be three main lines of objection.

a. Externalism: Proper Formulations

The first sort of objection may be traced to a possible misunderstanding concerning the nature of externalism as a theory of content. It takes the form of resisting what may seem to be the *prima facie* implausibility of an account of analytic sentences whose truth-conditions are made in the same terms as those of synthetic sentences. This kind of objection goes more or less as follows: suppose one takes Fregean senses to individuate mental states. Fregean senses are not in the head; indeed, Frege was adamant about their abstract nature. The view that takes Fregean senses to individuate mental states is not thereby, internalist, and - crucially - such a view preserves the classic distinction between sentences true in virtue of meaning alone and those true in virtue of meaning and the world. Such a view, it is claimed, is externalist at less cost (a related objection takes the favored individuating entity to be a structured proposition; the gist is the same).

Now I take it as obvious that it is an error to

suppose that internalism is circumvented by mere appeal to entities no one would claim are in the head to play the relevant role in the determination of a mental state. Something's being merely external to the head is insufficient to generate any interesting externalism. To see this, let us ask ourselves whether the following two accounts of the content of a belief differ in any crucial way.

One account specifies the content of

(7) Lois Lane believes that Superman is a hero

in terms of the belief relation holding between Lois and the structured *singular* proposition whose constituents are the individual particular Superman and the property of heroism. The other specifies the content of (7) in terms of the belief relation holding between Lois and the structured proposition whose constituents are, instead, the modes of presentation expressed by 'Superman' and by 'heroism,' where these do not involve the reference essentially. It seems obvious, in spite of the fact that both accounts specify belief as a relation to a proposition - an abstract object, one not in the head - that there is a difference in what it is that bears the individuating burden. As the controversy in the literature will attest, there is a critical difference in what each view takes to be the constituent

structure of the proposition, a difference that is essential to the account - and to the theoretical slant of the account - of the mental state in question. Clearly, it is insufficient to employ a vehicle of individuation that is not itself in the head in order to be able to claim that one's theory is an externalist one: it is quite consistent with *internalism* about meaning that a mental state is held to be characterized in the grasp of an abstract proposition. For suppose that one holds that the grasp of a proposition depends solely on internal facts about the subject of the mental state. Such a view would seem to be unarguably internalist. In order for a theory of the individuation of mental states to pass muster as externalist it is clear that an account both of the component structure of the vehicles of individuation, and of the role played by those vehicles with respect to certain conditions of understanding that is consistent with externalist requirements, must be supplied.

b. Mentalism about Properties Opposed

The second and third lines of objection are somewhat related. One is a denial that an account of analytic truth that claims that the truth-conditions of analytic sentences are to be made in property-revealing terms could be properly externalist, since properties are abstract objects, and as such, are world-independent. The other takes the form of the familiar claim that the relevant notion of a property is,

essentially, a semantic one. If this were so, it would naturally rob much of the interest from our view, since it would effectively collapse it into the standard view. If properties really are just semantic entities, then claiming that analytic sentences are true in virtue of necessary relations between properties would seem to be nothing but a verbal idiosyncrasy.

To deny that properties are of an essentially semantic nature is something of quite independent interest and deserves minute attention. Nevertheless, a number of basic - possibly familiar - remarks against a sense-property identity theory can be profitably made here.¹³ These will further aid in answering the objection against our account of analytic truth as properly externalist. Briefly:

(i) It seems implausible that there is no distinction between 'The carpet has the property of being red' and 'The carpet has the sense of the word 'red.' Objects instantiate properties, they do not instantiate senses.

(ii) It appears to be in dispute, as Fregean exegesis attests, whether or not senses depend on language, but it is again implausible that properties do. If properties are not distinguished from senses, some commentators are in the position of defending the view that properties depend on language. Surely, however, whether the carpet is red is a state of affairs quite independent of our linguistic faculty.

(iii) On the usual criterion for distinguishing sense from reference, one must consider whether there can be non-synonymous terms that have the same reference. This criterion applies no less to predicates than it does to singular terms. Both 'water' and 'H₂O' are predicates with the same reference, but they do not have the same sense.

(iv) Consider sentences that express identities between properties. It would seem that to claim that properties are senses would be to deny that there can be informative identities of this kind. For instance, one standard way to distinguish

(a) Water is water
from

(b) Water is H₂O

is on the grounds that (b) is a synthetic necessity. The properties denoted by the terms are identical but the sense of 'water' and that of 'H₂O' are not; hence, classically, (b) is informative but (a) is trifling. Suppose now that one holds that properties are senses. In both sentences the property denoted by the terms 'water' and 'H₂O' is the same. If properties are senses then the sense of 'water' is identical with the sense of 'H₂O'; not only that, the sense of 'water' is identical with the property of being H₂O, the reference of both terms. The implausibility of such a view

is shown, among other things, in that it surrenders the informativeness of (b) - not a happy result.

(v) Harking back for the moment to twin Earth, it seems that a sense-property identity theorist is committed to the following:

(1) that water is not H₂O. The fact that the sense of 'water' and that of 'H₂O' are not the same sense would render being water and being H₂O different properties, if properties are senses.

(2) that water is XYZ. The communities of Earth and twin Earth associate the same concepts with the word 'water.' The sense of both terms is thus the same; and if properties are senses, this would be sufficient to conclude, in spite of the fact that 'water' on twin Earth refers to XYZ, that water is XYZ.

(3) that H₂O is XYZ. Both 'water' and 'watert' have the same sense. But 'watert' is a word we've coined to translate what twin Earthians call 'water' so as to keep things straight for ourselves. Crucially, what it is that we are distinguishing is that 'water' and 'watert' do not have the same reference. A sense-property identity theorist seems hard pressed to account for this move.

c. Externalism and Abstracta

Now, while the foregoing considerations may help to break down resistance to rethinking a sense-property

identity theory, what is at issue, insofar as our account of analytic truth is concerned, is to claim a role for properties as entities of bona fide externalist individuation. Denying that they are essentially semantic entities is a start; but it might be further objected that it is their status as abstract objects that disqualifies them as the sort of individuating entities properly relevant to an externalist theory.

A few demarcations can serve as an indication of the line of argument we find compelling. First, we must find a plausible way to claim that a property, in spite of being an abstract object like a proposition or a Fregean sense, is not thereby a denizen of what some (Barwise and Perry 1983, 4) have called a "third realm" of abstraction. We have already noted that if one takes the relevant individuating context to be simply what is not in the head, then senses and abstract propositions *per se* will count as vehicles of externalist individuation. And we have already seen that it is imperative to give some account of the constituent structure of these entities, for the theoretical character of a theory of content depends in some measure on what it is that is taken to constitute the grasp of a sense or the component structure of a proposition. Now, properties belong to the extra-cranial environment in this wide sense. Is it possible to place them more locally?

A useful distinction to draw in order to achieve our

ends is one that distinguishes entities that figure in an account of mind, and whose characterization is intended to satisfy the demands of a psychological explanation, from entities that figure in an account of what we may call the scientific world, and whose characterization is intended to satisfy the demands of explanatory theories of this world. While it is true, of course, that theories about the events and state of affairs of the scientific world are expressed as a set of propositions, it is not the case that it is these abstracta that are the relevant objects of theories of the head-independent environment. Theories of this environment will take different objects from theories of mind, say, where propositions and other abstracta will figure as the relevant theoretical objects, as well as the vehicles through which the theory is expressed.

If we reflect a moment on the innovative nature of externalism as a theory of mind, the plausibility of such a distinction makes itself felt. The claim that the content of mind is shaped and essentially constituted by a reality distinct from, indifferent to, and independent of our bodily limits is not a lackluster claim. To say that properties individuate mental states might indeed fall rather flat as a theoretical claim if it were to turn out that properties are, for instance, essentially semantic entities with no claim to any interesting membership in the context beyond the cutaneous limits of a mind. It is in fact the essential

distinction between heads on the one hand, and the events and states of affairs - the objects and properties that are the context in which those heads navigate - that makes externalism provocative as a theory of mind.¹⁴

The literature regarding properties and their role in the explanation of what we have called the scientific world¹⁵ has introduced the notion of a property's causal role in explanatory theories of the world. Properties, on this view, have causal properties themselves in virtue of their necessary structural complexity. Thus, for example, one way to explain the events and states of affairs of the scientific world is to appeal to the causal laws that govern that world in terms of necessary relations between properties. Properties are understood to have causal powers in virtue of their structure and the necessary relations they bear to other properties in virtue of such structure. Given this picture of the extra-cranial environment, it seems clear that we can neither banish every abstract object to a "third realm," effectively removing it from that context we take as having an essential mind-individuating role; nor need we accept that every abstract object is *per se* on a par with every other as far as individuating theses of mind are concerned.

7. Conclusion

Given externalism, analytic sentences are true in

virtue of the world - true in virtue of the necessary structure and relations borne by the properties that are the reference of their terms - yet analytic all the same. A natural correspondence theory of truth is preserved, with the result that for analytic sentences, intension may be said to determine truth-value, not just extension. Truth-value is delivered through extension to intension; the world is not cut out.

We have seen that the characteristic modal and epistemic features of analytic sentences may be preserved by appeal to the notion of a property-revealing sense. Although their truth-conditions are made in the very same terms as those of synthetic sentences - world-involving terms - analytic sentences are distinguished by the way the senses expressed by their constituent terms reveal the componential structure and necessary relations borne by the properties by which they are individuated. Grasp of these property-revealing senses is sufficient for knowledge of the truth of these sentences: it is this feature that essentially distinguishes analytic sentences from synthetic sentences, both necessary and contingent.

In short, there is no opposition, on this view, between 'true in virtue of meaning' and 'true in virtue of the world' - the former is a special case of the latter. Analytic sentences express necessary truths as always; the difference is that the necessity is necessity *de re*.

Analytic truths are knowable *a priori*, as always; the difference is that what is grasped is not a truth about mind but a truth about the world. This is of course how it should be: analytic truth is, after all, *truth* in virtue of meaning. Analytic truth is necessary truth about the world, knowable *a priori*.

NOTES TO CHAPTER 3, Part One

1. See Quine (1980, 130-138).
2. To one degree or another, depending on what degree of externalism one espouses. More on this below.
3. In what follows we consider different accounts of the constituent structure of a proposition to examine the consequences of certain variations in the accounts. We understand Frege, familiarly, to take the sense of a sentence as the proposition it expresses: a proposition whose constituent structure consists in the senses expressed by the words in the sentence.
4. Barwise and Perry (1981, 387-403) also employ this particular segment to introduce their way of recovering semantic innocence. Our treatment differs in that the discussion is framed in terms of externalism about meaning in general, and extended to the question of analytic truth.
5. It is held to be knowable *a priori* and necessarily true.
6. Situation semantics is, of course, another variety.
7. Salmon and Soames (1988) contains a number of articles that are the best introduction to a semantic theory of this kind.
8. An objection might be raised here that a TDR meets Davidson's challenge to exorcise appeal to intensional entities in an account of the behavior of expressions in indirect discourse only trivially, if at all. For how is it possible to claim, as we wish to, that a Fregean theory need not make an in principle commitment to an independent realm of distinct semantic entities to serve as the reference of expressions in opaque contexts, when a TDR does not support

even the initial erection of the Fregean account? On a view that holds that there is nothing more to the meaning of an expression than its reference, it does not seem possible to claim that expressions in a that-clause shift their reference at all, let alone that the reference-shift they experience amounts to nothing untoward when the dust has settled. To meet this objection is not brief work; it raises, among other things, the issue of iterated belief contexts and I must defer elaboration to another paper. Suffice it to say that from the fact that an expression refers to nothing but its ordinary reference in and out of an opaque context, it does not follow that the *sense* expressed by the expression in and out of an opaque context remains the same.

9. A familiar complication that arises is, of course, that Lois may instantiate (1) and not (2):

(2) Lois Lane believes that Clark Kent is a hero

This is something to be explained by any theory that takes the meaning of an expression to be exhausted by its reference; fortunately, it is not the ins and outs of Direct Reference Semantics that concerns us here. The options are considered and perfected in, among others, a number of the papers in Salmon and Soames, 1988. The bibliography gives further directions.

NOTES TO CHAPTER 3, Part Two

1. See also Quine (1980, 9;130).

2. We are accustomed to thinking of the difference between analytic and synthetic sentences as a distinction between sentences whose truth-value is determined independently of the world and those whose truth-value is not. Now, we might understand 'true independent of the world' to mean two things: either necessarily true - true independent of the way things are in any world - or, alternatively, true in virtue of entities whose nature is distinct from the world. The former sense of 'true independent of the world,' however, is insufficient for a distinction between analytic and synthetic sentences, for as Kripke (1980) and Putnam (1975a, 215-71) have shown, we can countenance some sentences as necessary but *synthetic*. And the latter sense is thwarted by externalism about meaning. Clearly, if we claim that meaning is neither independent of nor isolated from the environment, it is doubtful that the truth-value of a sentence can be held to be determined in virtue of meaning alone, independently of how things stand in the world.

3. I have already acknowledged the controversy surrounding analyticity and I do assume throughout that analytic truths are necessary and knowable *apriori*. A brief comment on a notorious counterexample to the necessity condition might be useful, however. 'I am here now' would appear to be known *apriori* yet be contingently true: I might well have been somewhere else. Kaplan's distinction between character and content is useful in taking the sting out of this putative counterexample. As Kaplan puts it (1977, 71-2):

. . . a truth in the logic of demonstratives, like 'I am here now' need not be necessary . . . how can something be both logically true and thus *certain*, and *contingent* at the same time? . . . in the case of indexicals the answer is easy to see . . . the bearers of logical truth and of contingency are different entities . . .

it is the character that is logically true . . . but it is the content that is contingent or necessary.

We might represent the distinction as a distinction of scope; that is, from the fact that:

(1) Necessarily, every utterance of 'I am here now' is a true utterance.

it does not follow that:

(2) Every utterance of 'I am here now' is necessarily true.

4. See Kripke (1980, 39) for the working definition, and see Burge (1992, 9) for a useful exposition of the lack of consensus concerning the characterization of analyticity.

5. Familiar problems arise for an account of certain kinds of expressions when reference is taken to determine meaning: the problem of opacity is the most notorious. See Salmon and Soames (1988) for a useful introduction to so-called direct reference semantics. See also chapter 3, part one, where I argue that semantic innocence can be thought to characterize Fregean semantics - in spite of appearances.

6. Here I agree with Forbes (1989, 132). I should say that it was my misfortune to have worked out this account before I came across his book. I was no less elated (and somewhat relieved) to discover support for it there. This is not to say that I am confident that he would agree with everything I have claimed in this paper.

7. A distinction between proper and improper subset with respect to the constituent structure of properties will help to make this clear. We can understand some of the componential relations between properties to be improper, so to speak; to include the identity of properties as well as the relation of proper parts of properties.

8. See McDowell (1984, 283-94) for a profitable comparison. McDowell argues there for what I would call *object-involving* senses for singular terms. I extend the idea to predicates, and claim that while all terms express senses that are property-involving, some of those senses will be property-revealing.

9. The structural composition of a property is not to be confused with its essence.

10. Evans (1982) and McDowell (1977, 1984) were the first to suggest and defend a distinction between descriptive (so-called Fregean) sense and an alternative, so called *de re* sense. Briefly, the distinction turns on the relation between sense and its bearer. Frege argued that the sense of an expression could exist independently of the reference of the expression. Evans and McDowell deny this, defending a necessary relation between the existence and identity conditions of the reference of an expression and its sense.

11. One thing I take to result from this view of analyticity is the need for a reconsideration of the familiar categories of the *de dicto* and the *de re* as they apply to propositions. In particular, I believe that this view of analyticity recommends abandoning entirely the notion of the *de dicto* proposition. I pursue this in another paper.

12. I am grateful to Graeme Forbes for supplying me with this and the following example. He did also include: 'Every number has a successor,' but I believe that this one is possibly more controversial. One might argue that mathematics is the sort of area where discoveries take place, and that it is more akin to a discovery about the nature of numbers that they in fact have successors, than anything about the meaning of the word 'number.'

13. There are others. See Putnam (1975b, 429-40). Perhaps, in addition, the standard debate concerning intension and extension has resulted in an unwarranted conflation of the sense and reference of a predicate. Since extension is not sufficient to fix the meaning of a predicate, given non-synonymous predicates with the same extension, intensions were invoked to satisfy the demand for the meaning of a predicate - but intensions are commonly held to be the property or attribute denoted by a predicate. This may be a case where convenience sacrifices clarity - or even truth.

14. See McGinn (1989, chapter 1) for more on this point.

15. See, among others, Armstrong (1983, 1989), Dretske (1977, 248-68), Shoemaker (1984, 206-33; 234-60), and Tooley (1987).

REFERENCES

- Armstrong, D.M. 1978. *Nominalism and Realism: Universals and Scientific Realism*, Vol. I. Cambridge: Cambridge University Press.
- . 1978. *A Theory of Universals: Universals and Scientific Realism*, Vol. II. Cambridge: Cambridge University Press.
- . 1983. *What is a Law of Nature?* Cambridge: Cambridge University Press.
- . 1989. *A Combinatorial Theory of Possibility*. Cambridge: Cambridge University Press.
- Barnes, J. 1985. *The Complete Works of Aristotle*. Princeton: Princeton University Press.
- Barwise, J. and J. Perry. 1981. Semantic Innocence and Uncompromising Situations. *Midwest Studies in Philosophy VI*, ed. P. French, T.E. Uehling, Jr., and H.K. Wettstein. Minneapolis: University of Minnesota Press: 387-403.
- . 1983. *Situations and Attitudes*. Cambridge, Mass.: MIT Press.
- Blackburn, S. 1984. *Spreading the Word*. Oxford: Clarendon Press.
- Block, N. 1980. *Readings in the Philosophy of Psychology: Vols. I and II*. Cambridge, MA: Harvard University Press.
- Block, N. 1986. Advertisement for a semantics for

psychology. *Midwest Studies in Philosophy*, Vol. 10.,
ed. P. French. T. Uehling, Jr. and H. Wettstein.
Minneapolis: University of Minnesota Press: 615-78.

Boer, S. 1989. Neo-Fregean Thoughts. *Philosophical
Perspectives* 3., ed. J. Tomberlin. Atascadero, CA:
Ridgeview Publishing Company.

Bealer, G. 1993. Universals. *Journal of Philosophy* Vol. XC
(1): 5-38.

_____. 1982. *Quality and Concept*. Oxford: Clarendon
Press.

Budd, M. 1989. *Wittgenstein's Philosophy of Psychology*.
London: Routledge.

Burge, T. 1977. Belief De Re. *Journal of Philosophy* 74 (6):
338-62.

_____. 1978. Belief and synonymy. *Journal of Philosophy*
75: 119-38.

_____. 1979. Individualism and the Mental. *Midwest
Studies in Philosophy*, Vol. IV., ed. P. French, T.E.
Uehling, Jr., and H.K. Wettstein. Minneapolis:
University of Minnesota Press: 73-121.

_____. 1982. Other Bodies. *Thought and Object*, ed.
Andrew Woodfield. Oxford: Clarendon Press: 97-120.

_____. 1986. Individualism and Psychology.
Philosophical Review 95 (1): 3-46.

_____. 1988. Individualism and Self-Knowledge. *Journal
of Philosophy* 85: 649-63.

Carruthers, P. 1987. Russellian Thoughts. *Mind* 96, 18-35.

_____. 1988. More Faith than Hope: Russellian Thoughts

Attacked. *Analysis* 48: 91-96.

Churchland, P.S. 1986. *Neurophilosophy: Toward a Unified Science of the Mind/Brain*. Cambridge, MA: MIT Press.

Davidson, D. 1980. *Essays on Actions and Events*. Oxford: Clarendon Press.

_____. 1984. *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.

Davies, M. 1992. Aims and Claims of Externalist Arguments. Draft.

Dennett, D. 1988. Review of Psychosemantics. *Journal of Philosophy* 85: 384-89.

Donnellan, K. 1962. Necessity and Criteria. *Journal of Philosophy* LIX: 647-658.

_____. 1966. Reference and Definite Descriptions. *Philosophical Review* LXXV (3): 281-304.

Dretske, F. 1977. Laws of Nature. *Philosophy of Science* 44: 248-68.

_____. 1988. *Explaining Behavior: Reasons in A World of Causes*. Cambridge, MA: MIT Press.

Egan, F. 1991. Must Psychology Be Individualistic? *Philosophical Review* Vol.C (2): 179-203.

Evans, G. 1977. The Causal Theory of Names. *Naming, Necessity and Natural Kinds*, ed. S. Schwartz. Ithaca: Cornell University Press: 192-215.

_____. 1982. *The Varieties of Reference*. Oxford: Oxford University Press.

- Evnine, S. 1991. *Donald Davidson*. Oxford: Polity Press.
- Field, H. 1977. Logic, meaning and conceptual role. *Journal of Philosophy* 74: 379-408.
- _____. 1978. Mental Representations. *Readings in the Philosophy of Psychology*, Vol. 2, ed N. Block. Cambridge, MA: Harvard University Press: 78-114.
- Fodor, J. 1978. Propositional Attitudes. *The Monist* 61 (4): 501-23.
- _____. 1980. Methodological Solipsism Considered as a Research Strategy in Cognitive Science. *Behavioral and Brain Sciences* (3): 63-109.
- _____. 1981. *Representations: Philosophical Essays on the Foundations of Cognitive Science*. Cambridge, MA: MIT Press.
- _____. 1987. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.
- _____. 1990. *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press.
- _____. 1991. A Modal Argument for Narrow Content. *Journal of Philosophy* 88 (1): 5-26.
- _____. 1994. *Mentalese and Its Semantics: The Jean Nicod Lectures*, forthcoming.
- Forbes, G. 1989. *Languages of Possibility*. Oxford: Basil Blackwell.
- Frege, G. 1980. *Philosophical and Mathematical Correspondence*. Chicago: University of Chicago Press.
- _____. 1984. *Collected Papers on Mathematics, Logic, and Philosophy*. Oxford: Basil Blackwell.

Haugeland, J. 1980. Programs, Causal Powers, and Intentionality. *Behavioral and Brain Sciences* (3): 432-33.

Hempel, C. 1965. *Aspects of Scientific Explanation*. New York: Free Press.

_____. 1969. Reduction: Ontological and Linguistic Facets. *Philosophy, Science and Method: Essays in Honor of Ernest Nagel*. Eds. S. Morgenbesser, P. Suppes, and M. White. New York: St Martin's Press: 179-199.

Hume, D. 1978. *A Treatise of Human Nature*. Oxford: Clarendon Press.

Kant, Immanuel. 1965. *Critique of Pure Reason*. New York: St. Martin's Press.

Kaplan, D. 1968. Quantifying In. *Synthese* 19: 178-214.

_____. 1978. Dthat. *Syntax and Semantics 9: Pragmatics.*, ed. P. Cole. New York: Academic Press: 221-43.

_____. 1979. How to Russell a Frege-Church., in Loux, 1979.

_____. 1989. Demonstratives: an Essay on Semantics, Logic, Metaphysics and Epistemology of Demonstratives and other Indexicals. *Themes From Kaplan*, ed. J. Almog, J. Perry and H. Wettstein. Oxford: Oxford University Press: 481-563.

Kitcher, P. and W. Salmon. 1989. *Minnesota Studies in the Philosophy of Science. Vol. XIII: Scientific Explanation*. Minneapolis: University of Minnesota Press.

Kripke, S. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.

- _____. 1988. A Puzzle About Belief. *Propositions and Attitudes.*, eds. N. Salmon and S. Soames. Oxford: Oxford University Press: 102-48.
- Loar, B. 1980. Syntax, Functional Semantics and Referential Semantics. *Behavioral and Brain Sciences* 3: 89-90.
- _____. 1981. *Mind and Meaning*. Cambridge: Cambridge University Press.
- _____. 1987. Social Content and Psychological Content. *Contents of Thought.*, eds. R. Grimm and D. Merrill. Tucson: University of Arizona Press.
- Loewer, B. and G. Rey, eds. 1991. *Meaning in Mind: Fodor and His Critics*. Oxford: Basil Blackwell.
- Loux, M. 1979. *The Possible and the Actual: Readings in the Metaphysics of Modality*. Ithaca, NY: Cornell University Press.
- Marcus, R. B. 1971. Essential Attribution. *Journal of Philosophy* 68 (7): 187-202.
- Millikan, R. 1984. *Language, Thought and Other Biological Categories*. Cambridge, MA: MIT Press.
- _____. 1986. Thoughts Without Laws; Cognitive Science With Content. *Philosophical Review* 99: 232-53.
- _____. 1989. Biosemantics. *Journal of Philosophy* 86: 281-97.
- _____. 1991. Compare and Contrast Dretske, Fodor and Millikan on Teleosemantics. *Philosophical Topics* 18.2.
- McCulloch, G. 1988a. Faith, Hope and Charity: Russellian Thoughts Defended. *Analysis* 48: 84-90.
- _____. 1988b. Carruthers Repulsed. *Analysis* 48: 96-100.

- _____. 1989. *The Game of the Name: Introducing Logic, Language and Mind*. Oxford: Clarendon Press.
- McDowell, J. 1977. On the Sense and Reference of a Proper Name. *Mind* 86: 159-85.
- _____. 1984. *De Re Senses*. *Philosophical Quarterly* 136: 283-294.
- McDowell, J. and P. Pettit, 1986. *Subject, Thought and Context*. Oxford: Clarendon Press.
- McGinn, C. 1977. Charity, Interpretation and Belief. *Journal of Philosophy* 74: 521-35.
- _____. 1982. The Structure of Content. *Thought and Object*, ed. Andrew Woodfield. Oxford: Clarendon Press 207-58.
- _____. 1988. Review of *Psychosemantics*, in *Nature*, 331 (21 January): 615-619.
- _____. 1989. *Mental Content*. Oxford: Basil Blackwell.
- _____. 1991. *The Problem of Consciousness*. Oxford: Basil Blackwell.
- _____. 1991. Conceptual Causation. *Mind*, Vol C (October): 573-86.
- Nagel, E. *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt, Brace and World.
- Noonan, H. 1986. Russellian Thoughts and Methodological Solipsism. *Language, Mind and Logic* ed. J. Butterfield. Cambridge: Cambridge University Press: 67-90.
- _____ 1991. Object-Dependent Thoughts and Psychological

Redundancy. *Analysis* 51.1: 1-9.

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

Peacocke, C. 1981. Demonstrative Thought and Psychological Explanation. *Synthese* 49: 187-217.

_____. 1986. Reply to Stephen Schiffer. *Mind and Language* 1: 393-5.

_____. 1989. What Are Concepts? *Midwest Studies in Philosophy (Contemporary Perspectives in the Philosophy of Language II)*. Minneapolis: University of Minnesota Press: 1-28.

Pettit, P. and J. McDowell, eds. 1986. *Subject, Thought and Context*. Oxford: Oxford University Press.

Popper, K. 1972. *Objective Knowledge: An Evolutionary Approach*. Oxford: Clarendon Press.

Preti, C. 1992. Opacity, Belief and Analyticity. *Philosophical Studies* 66 (3): 297-306.

Putnam, H. 1973. Meaning and Reference. *Journal of Philosophy* LXX (19): 699-711.

_____. 1975a. The Meaning of 'Meaning.' *Mind, Language and Reality: Philosophical Papers, Vol. 2*. Cambridge: Cambridge University Press: 215-71.

_____. 1975b. The Nature of Mental States. *ibid.*: 429-40.

_____. 1975c. It Ain't Necessarily So. *Mathematics, Matter and Method: Philosophical Papers, Vol.1*: 237-49.

_____. 1975d Analyticity and apriority: beyond Wittgenstein and Quine. *Realism and Reason*:

Philosophical Papers, Vol. 3: 115-38.

_____. 1975e. Is Semantics Possible? *op.cit.*, Vol. 2: 139-152.

_____. 1975f. The 'Corroboration' of Theories. *op cit.*, Vol 1: 250-269.

_____. 1988. Synonymy and the Analysis of Belief Sentences. *Propositions and Attitudes*, eds. N. Salmon and S. Soames. Oxford: Oxford University Press: 149-58.

Pylyshyn, Z. 1984. *Computation and Cognition*. Cambridge, MA: MIT Press.

Quine, W.V.O. 1976. Carnap and Logical Truth. *Ways of Paradox and Other Essays*. Cambridge, MA: Harvard University Press: 107-32.

_____. 1976b. Quantifiers and Propositional Attitudes. *ibid.*: 185-96.

_____. 1980a. Two Dogmas of Empiricism. *From a Logical Point of View*. Cambridge, MA: Harvard University Press: 20-46.

_____. 1980b. Notes on the Theory of Reference. *ibid.*, Cambridge, MA: Harvard University Press: 130-38.

Richard, M. 1988. Direct Reference and Ascriptions of Belief. *Propositions and Attitudes*, eds. N. Salmon and S. Soames. Oxford: Oxford University Press: 169-96.

Russell B. 1912. Knowledge by Acquaintance and Knowledge by Description. *The Problems of Philosophy*. Oxford: Oxford University Press (1959): 46-59.

_____. 1918. The Philosophy of Logical Atomism. *Bertrand Russell: Logic and Knowledge*, ed. R. M. Marsh. London: Unwin-Hyman (1956): 175-283.

- Sainsbury, R.M. 1985. Critical Notice: Review of *The Varieties of Reference*. *Mind* 94: 120-142.
- Salmon, M. 1989. *Explanation in the Social Sciences*. Kitcher and Salmon, 1989.
- Salmon, N. 1982. *Reference and Essence*. Oxford: Basil Blackwell.
- _____. 1986. *Frege's Puzzle*. Cambridge, MA: MIT Press.
- Salmon, N. and S. Soames. eds. 1988. *Propositions and Attitudes*. Oxford: Oxford University Press.
- Salmon, W., et al. 1992. *Introduction to the Philosophy of Science*. Englewood Cliffs, NJ: Prentice-Hall.
- Schiffer, S. 1986. Peacocke on explanation in psychology. *Mind and Language* 1: 262-71.
- _____. 1991. Ceteris Paribus Laws. *Mind* C (1): 1-17.
- Segal, G. 1989a. Seeing What Isn't There. *Philosophical Review* 98: 189-214.
- _____. 1989b. The Return of the Individual. *Mind* 98: 39-57.
- _____. 1991. Defence of a Reasonable Individualism. *Mind*, Vol.C (4): 485-94.
- Shoemaker, S. 1984. *Identity, Cause and Mind: Philosophical Essays*. Cambridge: Cambridge University Press.
- Soames, S. 1988. Direct Reference, Propositional Attitudes, and Semantic Content. Salmon and Soames, 1988.
- Stampe, D. 1977. Toward a Causal Theory of Linguistic Representation. *Midwest Studies in Philosophy*, Vol. 2.

Minneapolis: University of Minnesota Press.

Stich, S. 1978. Autonomous Psychology and the Belief-Desire Thesis. *The Monist* 61 (4): 573-91.

_____. 1983. *From Folk Psychology to Cognitive Science*. Cambridge, MA: MIT Press.

Strawson, P.F. 1956. On Referring. *Essays in Conceptual Analysis*, ed. Anthony Flew. London: Macmillan: 21-52.

_____. 1959. *Individuals*. London: Methuen.

_____. 1974. *Subject and Predicate in Logic and Grammar*. London: Methuen.

Tooley, M. 1987. *Causation: A Realist Approach*. Oxford: Clarendon Press.

Villanueva, E. 1992. *Information, Semantics and Epistemology*. Oxford: Basil Blackwell.

Wittgenstein, L. 1969. *On Certainty*. New York: Harper.

_____. 1980. *Remarks on the Philosophy of Psychology. Vol I*: ed. G.E.M. Anscombe and G.H. von Wright. Oxford: Basil Blackwell.

_____. 1981. *Zettel*. Ed. G.E.M. Anscombe and G.H. von Wright. Oxford: Basil Blackwell.

Woodfield, A. 1982. *Thought and Object*. Oxford: Clarendon Press.

Yablo, S. 1992. Mental Causation. *Philosophical Review* 101 (2): 245-280.