

EXTERNALISM AND THE CAUSAL EFFICACY OF CONTENT

by

CARRIE FIGDOR

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

2005

UMI Number: 3159210

Copyright 2005 by
Figdor, Carrie

All rights reserved.

INFORMATION TO USERS

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 bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send 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.

UMI[®]

UMI Microform 3159210

Copyright 2005 by ProQuest Information and Learning Company.

All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Copyright

2005

CARRIE FIGDOR

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.

1/27/05

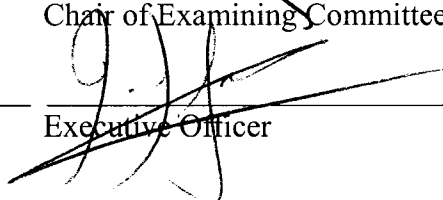
Date



Chair of Examining Committee

1/26/05

Date



Executive Officer

Michael Devitt

Martin Davies

Alberto Cordero

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

EXTERNALISM AND THE CAUSAL EFFICACY OF CONTENT

by

Carrie Figdor

Adviser: Professor Michael Devitt

This dissertation examines a *prima facie* conflict between two claims: first, that ordinary (or “folk”) psychological states play causal roles in psychology in virtue of their contents; and second, that their contents are extrinsic properties. This conflict, sometimes called the problem of externalism, is one of the difficulties facing any complete account of mental causation. I argue that there is no *a priori* reason why “wide” contents (as they are often called) do not or cannot play causal roles in psychological explanations of behavior, and show how they might do so.

In typical psychological explanations, beliefs, desires and other mental states are cited as causes of behavior: Socrates remained in an Athenian jail because he believed that it was unjust to disobey the city’s laws and he desired to act justly. If externalism is true, these contents – e.g. *that it is unjust to disobey the city’s laws* – are individuated partly in terms of the nature of the agent’s physical or social environment. It follows that if the environment is relevantly different, contents differ – even if there is no physical difference in the agent.

A variety of arguments try to show that externalism is not compatible with the causal role of content. I show how each argument fails to support its conclusion. I take one of Jerry Fodor's arguments as the most difficult. Fodor argues that wide contents violate the principle that mental properties must supervene on, or be identical to, neural properties if mental causation is to be made "intelligible". I show that his argument rests on the assumption that these neural properties are intrinsic, and argue that this "intrinsic hypothesis" is unjustified based on evidence from quantum chemistry and neuroscience. I also argue that at least some of the neural properties that play causal roles in computational theories of mind may be extrinsic. If so, then "wide" contents can be identified with or supervene on these properties, and they would satisfy Fodor's principle. I conclude by showing how this solution to the problem of externalism meshes with a general account of mental causation recently defended by Derk Pereboom.

Acknowledgements

My heartiest thanks go to Michael Devitt for his guidance throughout the development of this dissertation and the ideas that it contains. In his straightforward Australian manner, Michael refused to accept any line of argument that was not clear and well-supported. His resistance to my persistence stopped me from walking blithely over the edge of more than one intellectual cliff.

I also thank David Rosenthal and members of the Cognitive Science Symposium and Discussion group (in particular, stalwarts Jared Blank, Nada Gligorov, Pete Mandik, Doug Meehan, Roblin Meeks, Rebecca Robare and Josh Weisberg) for comments and questions raised during several presentations of dissertation-related material. John Greenwood provided valuable comments on early drafts, and he, Martin Davies and Alberto Cordero also deserve many thanks for their insightful comments on the final version.

I also thank Jeffry Ramsey, Andy Smith and Erik Banks for their help with quantum mechanics and physical chemistry, and Robert Francescotti for comments regarding the intrinsic/extrinsic distinction.

On a personal note, I am eternally grateful to Bill and Liz for their unflagging support during my entire *saison en enfer*; Svetlana Starkman, for making it all possible; Onyoung Oh, who helped me pull through many crises; Tom Sahagian, Esterlinda Keller, Johanna Stadler and the rest of the Swattie contingent for cheering me on; Tom Kent and Frank Crepeau, who made my transition from journalism easier than it might have been; Bruce Burger, Merav Brodetz, Sergio Gallegos and Pheroze Wadia, for their continual generosity and support; and my dear Guaicaipuro, who was always, completely there for me.

EXTERNALISM AND THE CAUSAL EFFICACY OF CONTENT

TABLE OF CONTENTS

CHAPTER ONE: SETTING THE STAGE

1.1 Introduction	1
1.2. Content: Representation, Individuation, Vehicles	6
1.2.1. Content and Intentional Content	7
1.2.2. Individuation of Content	11
1.2.3. Objects of Belief and Vehicles of Content	19
1.3. Mental Ontology	21
1.3.1. Essential Properties and Causal Powers	21
1.3.2. Mental Realisms	24
1.3.3. Descriptions and Identity Conditions	27
1.3.4. Events, Processes and States	28
1.4. Causation and Causal Explanation	31
1.4.1. Causal Relata and the Role of Properties	31
1.4.2. Regularity and Counterfactual Views of Causation	35
1.4.3. Causation and Causal Explanation	37
1.5. Extrinsic Properties	42
1.5.1. The Intrinsic/Extrinsic Distinction	43
1.5.2. Mere Cambridge and Functional Properties	57
1.6. Behavior	62
1.6.1. "Behavior"	63
1.6.2. Intentional Action	66
1.7. Related Topics	68
1.7.1. The Status of Folk Psychology	69
1.7.2. Non-Reductive Physicalism	70
1.7.3. Semantic Naturalism	72
1.7.4. Models of Cognition	75
1.8. Chapter Summary	77
CHAPTER TWO: SOME INITIAL ARGUMENTS	78
2.1. Introduction	78
2.1.1. Review	78
2.1.2. Extrinsic Causal Powers	79

2.2. Fodor's Argument from the Formality Condition	87
2.3. Arguments from Narrow Explananda	94
2.3.1. Stich's Replacement Argument	94
2.3.2. Kim's Causal Core Argument	103
2.4. Fodor's Cross-Context and Modal Arguments	115
2.4.1. Fodor's Cross-Context Argument	116
2.4.2. Fodor's Modal Argument	123
2.5. Chapter Summary	132
CHAPTER THREE: ARGUMENTS FROM NON-LOCALITY	135
3.1. Introduction	135
3.1.1. Review	135
3.1.2. General Remarks on Arguments from Non-Locality	137
3.2. Arguments from "Referential" Non-Locality	140
3.2.1. McGinn's Argument from "Referential" Non-Locality	141
3.2.2. Specific "Referential" Arguments	152
3.3. Arguments from Spatiotemporal Non-Locality	155
3.3.1. Fodor's Argument from Spatial Non-Locality	155
3.3.2. Dretske's Argument from Temporal Non-Locality	182
3.3.3. Summary of Non-Locality Arguments	188
3.4. One Last Argument, and On Being A Planet	189
3.4.1. Fodor's 'Technical' Argument	189
3.4.2. On Being a Planet	190
3.5. Chapter Summary	195
CHAPTER FOUR: AN ACCOUNT OF EXTRINSIC CAUSAL POWERS	197
4.1. Introduction	197
4.2. Extrinsic Causal Powers in the Physical Sciences	201
4.2.1. The Extrinsicness of Molecular Structure	201
4.2.2. The Resistance to Extrinsic Causal Powers	218
4.3. Wide Semantic Causal Powers	222
4.3.1. Semantic Structural Properties	223
4.3.2. Robust Mental Causal Powers	234
4.4. Chapter and Dissertation Summary	245

Appendix I: Woolley's Argument in More Technical Terms	247
Bibliography	257

CHAPTER ONE: SETTING THE STAGE

1.1. INTRODUCTION

The topic of this dissertation is a special case of a long-standing philosophical problem: How can the mind cause anything? The general problem of mental causation is to explain how mental features can play a causal role in a world in which a physical causal explanation can be given for any singular event. The problem would not arise if we did not think mental states could cause anything, or if we did not think there were any mental states, or if we did not hold that physical causal explanations were universally available.

But the existence and causal efficacy of the mental is something that most of us are deeply committed to. As Socrates pointed out, it is one thing to say that I am in an Athenian jail because my bones and sinews moved in certain physically explainable ways, and quite another to say that I am in jail because I believe that the laws of the State have condemned me and I have decided to obey those laws.¹ The truth of certain counterfactuals – “If I had decided otherwise, I would have fled to Megara long ago” -- expresses the commitment to mental causation that is presupposed in human agency and in a psychology that explains behavior by reference to mental phenomena. We are also committed to mental causation in any theory of human knowledge in which perceptual states have physical causes and mental effects, and in which these mental effects – beliefs and memories, for example – interact in theoretical and practical reasoning.² These

¹ Plato, *Phaedo* (98c-99a).

² Kim (1998: 29-56) distinguishes three ways in which we are committed to mental causation: it is presupposed in the possibility of human agency, the possibility of human knowledge and the possibility of a scientific psychology.

commitments motivate an explanation of mental causation because denying their reality is, for most of us, a bullet that is far too big to bite.

The particular version of this problem that I am interested in involves that class of mental causes whose mental features are dependent in part on the world beyond the agent. Members of this class are characterized partly in terms of how things are in the agent's environment, or her relations to that environment. Such mental states are typically said to be "externalistically", "externally", "widely", "extrinsically" or "relationally" individuated. Explaining how mental phenomena characterized by such relational features can play a role in causal relations may be called the problem of externalism. The problem is to understand how something's being an extrinsic property of an entity can play a role in the causal relations in which that entity takes part. As an analogy, and intuitively, the fact that a particular billiard ball was made in Tennessee – a relational, and historical, property of the ball – seems irrelevant to its ability (given an impulse by a cue stick) to knock another billiard ball into the corner pocket; that ability is a matter of the ball's current, internal features. Similarly, assuming that a belief about Athenian law is a neural state with the appropriate causal-historical relations to Athenian society, the fact that a particular neural state has these relational features seems irrelevant to its power to cause activation in motor-output neurons; intuitively, its activity is just a matter of its current, internal features. In short, causation seems to be a matter of relations between synchronic, internal features of particulars, yet many important mental features are relational (or extrinsic) and historical.

The problem of externalism is especially pressing because the class of externally individuated mental states includes beliefs, desires and other psychological states that

have long played, and continue to play, an indispensable role in reliable and common causal explanations of human behavior. A case in point is our explanation of Socrates' presence in jail in terms of his beliefs and desires. If an adequate account of the causal-explanatory role of such states as beliefs and desires cannot be given, then this everyday explanatory framework has little or no justification.

It might be argued that the above is not sufficient motivation for a dissertation on wide semantic causal powers. The intuitions about causation that motivate the problem are themselves part of an everyday conception of physical causation and explanation. But relational properties are found all over the sciences, including contemporary physics. So why shouldn't pointing out this fact be sufficient to silence critics of wide semantic causal powers? The first reason is that it *hasn't* been sufficient; Fodor, for example, is well aware of extrinsic causal powers in science, but proceeds to argue against semantic causal powers anyway. The second reason explains the first: pointing out this fact does nothing to shake the conviction that molecularly identical brains (Twin cases) would operate in exactly the same way even if they realized distinct types of thoughts. It takes some work to show exactly how this conviction departs from science. No general observation about the role of extrinsic properties in science can do this work. I try to do it in what follows.

Thus, I have two goals in this dissertation: first, to examine arguments that purport to show that the relational features of mental states, in particular the contents of our ordinary beliefs and desires, do not or cannot play any genuinely causal roles; and second, to suggest in broad terms a metaphysical framework, consistent with

contemporary science, which allows for the possibility that they can play genuine causal roles given the doubts that have been raised about them.

In the first case, the only problems I examine are those in the literature that raise issues for mental causation that can be traced to the relational nature of some mental states; and the only question I address is whether or not these arguments establish that there are genuine problems for mental causation arising from their relational nature that are in addition to other standing problems of mental causation. The following illustrates the difference. Suppose two particulars x and y are causally related iff there is a law relating A 's and B 's such that x 's being A is appropriately related to y 's being B . This schema – the nomological view of causation – is neutral as to what sorts of laws there are and what kinds of properties can be substituted for ' A ' and ' B '. If Davidson is right that there are no psychological laws, no psychological properties of any kind can be substituted for ' A ' and ' B '. But if relational properties in particular cannot play causal roles, then even if there *were* psychological laws, relational properties could not be substituted for ' A ' and ' B ' *anyway*. So the question I will examine is whether the arguments provided in the literature give us genuine reasons to think that this *latter* claim is true.

I will argue that they do not. I will show that there are no metaphysical reasons for barring relational properties from being causal powers, and that the epistemological problems that have been raised – in particular, that laws that subsume phenomena under relational properties will be insufficiently general or unscientific – are invalid. My conclusions leave untouched the other significant problems for mental causation. These include most prominently the problem of anomalous monism, which involves denying

that there are psychological or psychophysical laws, and the problem of causal exclusion (or of the supervenience of the mental), in which properties (such as mental properties) that supervene on physical properties are excluded or preempted by their physical supervenience bases from playing genuine causal roles.³ If there are no psychological or psychophysical laws, then given a nomological account of causation (in which a property is causally efficacious only if it appears in a causal law) no psychological properties, wide or otherwise, can be causal. If we suppose that an externalistic mental state supervenes on an agent (or on the agent plus her environment), then showing how this state can play a causal role in producing the agent's behavior in virtue of its externalistic character will not show how it can play a causal role in spite of its supervenient character. These other problems have spawned vast literatures that I must set aside for the purposes of this dissertation.

The question I address in pursuing my second goal is whether there is a metaphysical framework in which extrinsic mental properties could play genuine causal roles *if* these more general problems for mental causation were solved. I argue that there is – although I also claim that whether or not contents (in particular) are causal powers is ultimately a matter of empirical research. I first show how there can be structural properties that are extrinsic, examining the case of molecular structure; and then argue that some of the structural properties now included under the term “syntax” (in the computational theory of mind) might include semantic properties. Since “syntax” is widely accepted as causally efficacious, the problem of externalism dissolves if this

³ See Kim (op. cit) for detail on these problems. Kim has been a major figure in discussing the exclusion problem, while the problem of anomalous monism stems from Davidson's “Mental Events”, reprinted in his (1980). In a neutral formulation, supervenient properties are those that stand in a necessary correlation with physical properties, such that if the physical properties of an individual or a world are fixed, then that individual's or world's nonphysical properties are fixed.

account is correct. I also find a role for structural semantic properties in a general defense of robust mental causal powers recently proposed by Derk Pereboom. The positive thesis does not presuppose any particular solution to the other problems of mental causation.

In this chapter I will clarify my topic by explaining the technical philosophical vocabulary in which it is usually discussed and distinguishing this topic from related philosophical issues. The main concepts to be defined include what it is for a mental state to be about something, what it is for a mental feature to depend on external relations, and what a causal relation is. The basic conceptual repertoire is drawn from philosophy of language, philosophy of mind and metaphysics. It includes the concepts of intentional content and externalistic individuation, discussed in section 1.2; mental ontology, in 1.3; causation and causal explanation, in 1.4; relational properties, in 1.5; and behavior, in 1.6. In some sections I will use concepts explained in later sections.

For reference, closely related topics are briefly introduced in 1.7. These include the status of folk psychology, the viability of non-reductive physicalism, the project of naturalizing semantics, and the development of competing theories of cognition. I close with a brief restatement of the topic in 1.8. I will refer to sections of the text by chapter, section and subsection number; for example, “2.3.2” refers to chapter 2, section 3, subsection 2 and “2.3” refers to chapter 2, section 3.

1.2 CONTENT: REPRESENTATION, INDIVIDUATION, VEHICLES

Thoughts – including, for example, beliefs and desires – are one of the basic types of mental phenomena in our everyday psychological-explanatory repertoire. In current philosophical terminology, thoughts are mental states with intentional or representational

content. If thoughts play a causal role, it is in virtue of their content that they do: it is because Socrates desired *to obey Athenian law*, rather than *to escape the clutches of Athenian law*, that his presence in Athenian jail is explained. So the concept of content is basic to my topic. This section focuses on clarifying the concepts of intentional content, externalistic individuation of intentional content and the structure of contentful states. In 1.7.3, on semantic naturalism, I mention prominent theories of intentional content. I do not presuppose any particular theory of content.

1.2.1. Content and Intentional Content

As currently used, the term “content” is a general name applied to that feature of mental phenomena in virtue of which they count as mental. It is used to classify both phenomena with intentional content, such as beliefs and desires, and those with phenomenal content, such as feelings of pain or experiences of seeing red (assuming, at least for the sake of exposition, that this is a valid distinction between mental phenomena). While different philosophers define and label types of content differently, very roughly speaking the types of content that belong in the former category are often called semantic, representational, informational, conceptual or propositional contents, while contents in the latter class would include qualia or sensory or perceptual or nonconceptual content. All mental states have one or another type of content (or perhaps both) in this quite broad sense of the word “content”.

The term “content” has thus come to be used for a wide variety of mental phenomena in much the way Brentano used the term “intentional” as the mark of the mental. However, as defined by Brentano, “intentional” denoted just the feature of

“aboutness” in virtue of which mental states are directed towards, or represent, things other than themselves.⁴ This feature of “aboutness” is most clearly and closely associated with those mental states that have semantic content – those which represent or contain information about objects and situations in the world. However, as Brentano emphasized, it also characterizes mental states that are about non-existents (such as beliefs about Santa Claus).⁵ This feature of “aboutness” is also increasingly used to characterize basic cognitive processes such as vision that are a focus of research in cognitive neuroscience.⁶

As such, the phrase “representational content” (or perhaps “Brentano-intentional content”) has wider scope than that of what might be called “Chisholm-intentional content. In philosophy of language, a linguistic context is “intensional” if the expressions in it (i) do not allow for truth-value-preserving substitution of co-referring expressions or (ii) do not permit straightforward existential generalization (they cannot be “quantified into”). Linguistic contexts whose expressions allow these operations are “extensional”. ‘That’-clauses are paradigmatically intensional contexts, for terms in the sentences appearing in them (e.g. “Santa Claus” in “John believes that Santa Claus is a babe magnet”) fail these tests. The mental states with which I will be concerned – the “propositional attitudes” – are typically picked out by the use of “that”-clauses. Relying on this feature of how mental states are picked out, Chisholm proposed a linguistic criterion of intentionality, based on the intensionality of that-clauses, to identify mental

⁴ In this section I generally follow Bechtel’s (1988: 40–47) discussion of Brentano’s and Chisholm’s differing views of intentionality.

⁵ Dretske (1995) and Rey (1997), among others, also explain the content of sensory states in terms of representation. These views, while they would bring phenomenal states within the scope of Brentano’s definition, remain controversial. I set such views aside since if they succeed, it would only extend the scope of “intentional state” to include states that are not normally included in the problem of externalism; and if they fail, the problem of externalism remains for the states that are already classified as “intentional”.

⁶ Egan (1994: 258), for example, characterizes visual states and processes as having “representational content”, and calls “intentional” any theory that taxonomizes them by their “representational content”.

phenomena. By this criterion, we distinguish phenomena with intentional content from phenomena that lack this feature by means of a special intensional vocabulary. The idea is to avoid Brentano's problem of explaining how thoughts can be "about" non-existents by drawing a mental-physical distinction in terms of differences in the language we use to pick out items in each category.

The fundamental difficulty with Chisholm's criterion is that there is no logical connection between the representational features of mental states and the logical features of linguistic expressions used to pick them out.⁷ How mental states represent the world and how our linguistic means for referring to those states operate are distinct issues.

I will use the term "intentional content" or simply "content" in the "Brentano-intentional" or representational sense, although in most cases with reference to a narrower class of states than Brentano allows: intentional or representational content will denote that feature of certain mental states in virtue of which they are about existing objects and features in the world. This usage will largely sidestep (leave unaddressed) the problem of accounting for how mental states can be about non-existents; for example, beliefs about Santa Claus would not be included in this class. Presumably any adequate externalist theory of content will have to provide or be compatible with an explanation of such beliefs.⁸ Whether this usage also leaves out states with sensory content will depend on one's account of sensory content. However, my usage clearly includes the contents of states that paradigmatically are representational and are individuated in part by reference to external existents. These are among the states that play the indispensable role in psychological explanations of behavior that needs to be accounted for.

⁷ Bechtel (op. cit) cites Searle among those who raise this objection to Chisholm's criterion.

⁸ However, the issue of reference to non-existents is raised by some philosophers (including Fodor and Pylyshyn) as a reason against considering intentional contents as causal. I discuss this issue in 3.3.2.

This sense of “intentional content” is built into the currently dominant view of cognition as involving *mental representations*.⁹ The model for mental representations is Russell’s characterization of mental states as “propositional attitudes” – that is, as mental attitudes (such as believing, desiring, hoping, and so forth) to non-mental propositions (which, by tradition, are abstract objects that encapsulate the truth-conditional meanings expressed by sentences and are denoted by means of that-clauses). Due to basically Quinean animosity toward the ontological excess of positing abstract objects, many philosophers eliminate the reference to propositions by positing sentences in a “language of thought” that is realized in the brain. However, the same basic feature of mentality emphasized by Brentano remains, only now it is thought of as that feature in virtue of which a Mentalese sentence represents the world. The term “mental representation”, rather than Mentalese sentence, is most commonly used, perhaps because it does not rule out mental images or connectionist models of the mind (see 1.7.4) that posit mental representations that are not sentences. In this standard framework, which I will assume, contents are semantic properties of mental representations. I also adopt the usual accompanying assumption that these semantic properties can be explained naturalistically – that is, materialistically and empirically in terms of non-semantic properties and relations (see 1.7.3 below on semantic naturalism).

Clearly much psychological explanation involves reference to mental states that are not typically identified in terms of representational content, such as emotions or sensations. I am restricting my attention to states with representational content for three reasons. First, the philosophical literature on psychological causal explanation focuses

⁹ Field (1978) presents the basic model, while Fodor (1975) presents a computational model of mind in which the manipulated entities are syntactically structured, semantically evaluable representations that form a language.

almost exclusively on the role of beliefs and desires; beliefs are usually taken as the paradigm representational mental state. A typical explanation of why we do what we do almost invariably and at a minimum involves reference to some combination of beliefs and desires. So any adequate account of psychological explanation must consider the role of these representational states. Second, accounting for the role of beliefs and desires is already quite complicated. So I am adopting a simplifying assumption made by almost all writers on this and related topics. Third, it is controversial whether emotions or sensations have content in the sense of interest here, let alone whether those contents are externally individuated. Since my focus is specifically on externally individuated contents, these matters would have to be settled before emotions and sensations could be determinately within the scope of my dissertation. So I will adopt the usual, if artificial, practice of ignoring their role in causing behavior.

1.2.2. Individuation of Content

Individuation conditions specify the existence and identity conditions of entities; they are the conditions under which an entity exists (or ceases to) in a possible situation and what sorts of changes an entity can undergo that preserve its identity. I will discuss particulars (or tokens) and their existence and identity conditions in 1.3 on mental ontology, and focus on properties (or types) and their identity conditions here (leaving open whether properties exist).

Assuming the standard model of the structure of intentional states (see 1.2.3), states with intentional or representational content can differ at the type level either by their content or their state or both. The belief *that the elephant in my bathtub is interested*

in getting out differs in content from the belief *that the punch has been spiked with rubbing alcohol*, differs in state-type from the *hope* that the elephant in my bathtub is interested in getting out, and differs in both from the *hope that the punch has been spiked with rubbing alcohol*. (At the token level, states are numerically distinct for numerically distinct agents or times.)

The individuation conditions of intentional contents (that is, of the semantic properties of mental representations) have received more attention than those of types of state, perhaps because contents play the larger role in explanation. However, the latter play a role as well: she opened the window because she hoped that the mosquito would fly out, not because she believed that it would. State-type individuation conditions are typically thought of in terms of differences of functional role in a mental economy (see 1.2.3 below for the “belief box” metaphor). These roles are assumed to mirror the “attitudes” of the original propositional-attitude model: belief, desire, hope, intention and so forth. It is an open question whether these categories will figure strictly, approximately or at all in a future intentional psychology. For example, a psychofunctionalist, such as Lycan (1988), may hold that the term “belief” picks out a psychological natural kind that a mature psychology will explain, while an eliminativist, such as Stich (1983), may hold that “belief” is referentially empty. I will remain neutral on this issue. However, since these states are the ones that play a role in ordinary psychological explanations of behavior, the categories of belief, desire and so forth are at least pragmatically useful. Therefore they stand at least *some* chance of being incorporated, perhaps in somewhat revised form, into a future intentional psychology (should we want one).

Individuation of content types is somewhat trickier. There are two main issues here: externalistic individuation and the relation between content and meaning. The latter issue is whether the individuation of the semantic properties of mental representations is or ought to be very closely linked to the individuation of the semantic properties or meanings of the content sentences in the that-clauses typically used to denote (or “ascribe”) contentful states. Since this issue will not play a role in my dissertation, I will leave open the question of how mental content and linguistic meaning are related.

The fundamental idea of externalistic individuation is that the contents of many intentional states are determined at least in part by the agent’s external relations to her physical and social environments.¹⁰ Contents (at least, those that are Mentalese sentences) are standardly thought of as being composed of concepts; my belief *that the woman in the ground-floor apartment has a new dog* contains the concept DOG. The complex issue of concept individuation will not be addressed here, although the truth of externalism entails that concepts, as semantic units in contents, are individuated at least in part externalistically.

Externalism claims, intuitively, that a token belief about dogs counts as a belief of that content-type (i.e. a dog-belief) only if, and partly because, the agent is in an environment that contains dogs; if there were no dogs for her or her fellows to be in contact with, neither she nor her fellows could have dog-thoughts.¹¹ More technically, if externalism is true, content-types differ if there is a difference in the kinds or natures of the entities in the subject’s environment which the constituent concept-types represent

¹⁰ Devitt (1996: 161) classifies these noninferential, causal links into three types: (1) historical causes, (2) reliable causes and (3) teleological causes. Externalistic contents can involve any or all of these.

¹¹ We can of course have thoughts about theoretical or rare entities – griffins, rare elements and such – but it is an open question whether such contents must be externalistically explained, as opposed to merely being explained in a way that is compatible with externalism. See footnote 8 above.

(see 1.3.1). As Burge (1982) puts it, the individuation conditions of ordinary contents are such that they “make reference to the nature of the entities in [an individual’s] environment, or at least to what his fellows consider to be the nature of those entities.”¹²

In contrast to externalistic individuation, individualistically (or “narrowly”) individuated content-types can be identical despite differences in the essential kinds of the entities to which the contents refer in a given context. Whatever these “narrow” contents have in common, it is not determined by the properties shared by the external entities in the agent’s environment. Also, the truth of externalism does not require that all contents be externalistically individuated, but that at least some are. For this reason Burge (1986) also calls the view “anti-individualism”.

The allegedly problematic aspect of externalistic individuation is this fact that it depends in part on the natures of things which happen to be instantiated in the agent’s world. If externalism is true, what is in an agent’s mind depends essentially on what is in her world because her contents essentially depend on which kinds are instantiated there (see 1.5). If there are no instances of doghood in an agent’s environment, either the thoughts we ascribe to her using the term “dog” are narrowly individuated or they are not about instances of doghood. Externalism leaves room for some, even many, contents to be explained in terms of indirect links to the environment. For example, unicorn-thoughts may be explained in terms of internally determined associations between externally determined horse- and horn-thoughts. But at the basic level at least there must be contents that are directly determined by such external relations.

¹² Similarly, Pettit and McDowell (1986: 4) characterize externalism as the idea that mental states are “world-involving” in that they “partly depend on the actual scientifically discoverable nature of something in the external world.”

Although I assume externalism is correct for the purposes of this dissertation, two main thought experiments have convinced many that the contents of many of our ordinary intentional states are at least in part externalistically individuated. One is Putnam's Twin-Earth thought experiment, which was initially introduced for the case of the meanings of linguistic expressions and has been extended to apply to the contents of mental states; the other was suggested by Burge.¹³ Both arguments are widely familiar, so my review of them will be brief.

In Putnam's thought experiment, two physically identical subjects – Oscar and Twin-Oscar – speak identically-sounding languages, have identical physically-described experiences and live on planets that are physically identical but for one thing: where there is H₂O on Earth, there's X_YZ on Twin-Earth. This other stuff is indistinguishable from water in every respect except for its chemical composition – that is, except for its nature. It is reasonable to think that when Oscar has a thought about the stuff he calls “water” he is thinking about H₂O, and when Twin-Oscar has a thought about the stuff he calls “water” he is thinking about XYZ. Conversely, there seems no reason to ascribe a belief about XYZ to Oscar, who has never been in contact with any (nor have his fellows); ditto for a belief about H₂O and Twin-Oscar. If these intuitions are sound, it follows that their content-types differ when the “natural kind” types instanced in their environments differ, even if their internal physical states and processes are identical.

Burge's original thought-experiment involved properties that are essential but are not intuitively “natural kinds”, where “natural kinds” are, roughly, those that correspond to simple predicates that appear in a science's laws. In his story, a nameless patient of a

¹³ See “The Meaning of ‘Meaning’ in Putnam (1975) and Burge (1979) for his original thought experiment. Burge has since presented several related arguments in other articles.

certain age complains to his doctor that he has arthritis in his thigh. His other beliefs about arthritis are largely true: he has it in his joints, it is less dangerous than liver cancer, there are certain aches that are characteristic of arthritis, and so on. It seems reasonable to suppose that in this case the Doctor corrects the Patient; that is, we ascribe to the Patient a false belief about arthritis, which is nevertheless a belief about arthritis and not about something else. We then suppose, counterfactually, that the Patient lives from birth through a series of events physically identical to those he actually lived through. However, in this counterfactual situation, the predicate “is arthritis” is used by the Patient’s linguistic community to refer to a condition that is instanced in thighs as well as joints. In this situation, it seems reasonable to ascribe to the Patient a true belief about an instance of this other condition (usually dubbed “tharthritis”). That is, even if all internal physical states and processes are identical, the Patient’s social setting determines which kind his uses of “arthritis” refers to; it does not refer to instances of the same kind as our term “arthritis”.

I should note that externalism is compatible with either realism or anti-realism in that in principle the natures of the entities to which externalistic states are related can be mind-independent or mind-dependent.¹⁴ However, I will assume realism. Thus, we (i.e. the social context) determine which kinds we (you and I) are thinking about, but we (the social context) do not determine the natures of those kinds. They have their natures independently of us and our names for them. (*How* we determine that “arthritis” refers to arthritis is the task of a theory of content; see 1.7.3 below.)

Burge’s thought-experiment may be considered a perspicuous generalization of Putnam’s. For example, it is a decision of our scientific community to count the isotopes

¹⁴ See Devitt (1997: 235-58) on realism and anti-realism.

of oxygen as variants of oxygen, not as different elements; in consequence we use the same name for them all. For a realist this only means that our scientists rank mind-independent differences between isotopes as taxonomically insignificant (for scientific explanatory purposes) relative to their similarities; it does not imply a socially-constructed world. In another world, scientists might find these isotopic differences important for explanation, and use a different name for each isotope. In this case, “oxygen” there would not have the same extension as it does here; it would be like “tharthrititis”. Similarly, if the nature of the stuff called “water” on Twin-Earth was not XYZ, but simply being a clear liquid (such that gin, XYZ, vodka, H₂O if there were any, etc. are all instances of “water”), then Putnam’s thought-experiment would be equivalent to Burge’s, albeit with Twin-Earth standing for a counterfactual case. This would not make being a clear liquid any less of a mind-independent property; it’s just that being a clear liquid is not a property that plays a role in our scientific explanations.

For my purposes, the distinction in McGinn (1989) between weak externalism (where the external entities can be properties existing in Plato’s Heaven) and strong externalism (where there must exist appropriate property instances somewhere in the spatiotemporal vicinity, or at least the universe, of the subject) does not matter; the worry about externalistic properties in this context is that they’re dependent on something outside the agent, and this worry applies equally to weak and strong externalism.¹⁵ However, I will largely assume strong externalism, since that is the usual conception and since everyone can accept cases of externalistic states involving properties instantiated in the agent’s environment without making further metaphysical commitments about

¹⁵ McGinn (1989: 131) himself agrees that the problem of external causal powers that he raises applies in either case. I discuss his argument in Chapter 3.

properties. Moreover, strongly externalistically individuated states would seem to play more important roles in explaining behavior.

I also note that externalism about content is distinct from but compatible with externalism about the vehicles of content suggested by Hurley (1998) – also called the “extended mind” thesis, discussed in Chapter 3 – whereby the (presumably physical) contentful entities may have parts beyond the agent’s body.¹⁶

A special case (or closely related group of cases) of externalistic mental states are what have been called demonstrative beliefs, indexical beliefs, or (in one use of “de re”) de re beliefs. According to Burge (1977), a de re belief in this sense “is a belief whose correct ascription places the believer in an appropriate nonconceptual, contextual relation to objects the belief is about.” What makes these beliefs externalistic is that it is not possible to have a token belief of this sort unless the agent stands in an external relation to an object at a time, whether or not any further (conceptual) condition is required or presupposed. What makes them a special case is that (1) the external relation is typically a spatiotemporally unmediated perceptual one (my belief that the telephone is ringing, when I am listening to it ring), as opposed to one that is socially or historically mediated, and (2) the nature of the object perceived is not explicitly characterized by the demonstrative used to help specify the content (e.g. my belief that *this* is ringing).¹⁷ These differences do not alter the basic fact that an external relation is required to have such beliefs, which is the issue of interest here. So I will ignore this special category of externalistic beliefs.

¹⁶ This claim will be defended at length in Chapter 3, when I discuss McGinn’s Non-Locality argument.

¹⁷ That is, to express a belief using “this is ringing” does *not* imply that the nature of the referent of “this” is irrelevant to determining the content of the belief, even though “this” does not specify that nature. This ensures that the general definition of externalistic contents -- as those whose individuation conditions make reference to the *nature* of the entities in environment -- is not violated by de re (demonstrative) beliefs.

Most philosophers today accept that the externalistic individuation of content is true for many, if not all, ordinary mental states. Burge has also argued that externalism is also true for the individuation of at least some mental states posited in cognitive psychology, in particular in the theory of vision.¹⁸ If so, then even if *folk* psychology were entirely discredited and the intentional states (like belief) that it posits were abandoned, the causal efficacy of externalistically individuated content would remain an issue for any future intentional psychology or cognitive science.¹⁹ Such states would not be intentional in Chisholm's sense, but they would be in the Brentano-intentional sense I am using. This possibility is an important topic in its own right, so I only note that the way I have defined representational content leaves open the possibility that my discussion could in principle be extended to cover non-folk-psychological intentional states.

1.2.3. Objects of Belief and Vehicles of Content

Ordinary representational states are also called propositional attitudes because entertaining a thought has traditionally been analyzed as the holding of a relation between a proposition and a subject. On this original relational view, believing is a complex state of a person in which the person has a certain attitude towards, or stands in a certain relation to, a particular proposition. Since belief is the paradigmatic propositional attitude in terms of which much of this debate is conducted, I will focus here on accounts of belief and make the usual assumption that the other propositional attitudes follow suit.

¹⁸ Defenses of individualism in response to Burge (1986) include Segal (1989) and Egan (op. cit.).

¹⁹ So *pace* Fodor (1987: 30), even if we don't end up with the identity conditions for intentional states that common sense prefers, the identity conditions we end up with in a scientific psychology may still be relational, and any concerns about the causal efficacy of relational properties will remain.

A reluctance to countenance the existence of abstract entities such as propositions, the desire for a philosophical account of content that is compatible with physicalism or at least naturalism, and the rise of computational theories of mind have led to a modification of the original view. The current orthodoxy, in its basic structure due to Field (1978), holds that believing is a naturalistically explainable, dispositional relation between an agent and a mental representation, such that *x believes that p* iff *x* is disposed to use tokens of that representational type in a certain way in cognitive operations. A mental representation is (on the leading, Fodorian, view) a syntactically structured sentence in a rule-governed “language of thought” (see 1.7.4); it is a symbol with content. Thinking, in turn, is analyzed in terms of computational relations between mental representations. A computational relation is a causal relation between symbols in virtue of their “form” (or “syntax”; this will be discussed further in chapter 4).

This basic model of what it is to believe something is often expressed in terms of a mental sentence being tokened in a “belief” box (or “desire” box, and so on for each attitude), where the spatial metaphor of an object’s being placed or appearing in a box stands for a mental representation’s functional “place”, or role, in a network of causal relations in a mental economy. Believing is using a mental representation in the way characteristic of that type of functional role (or some portion of it). I will assume this standard view of mental state-types (belief, desire and so on). The basic model leaves open what the “vehicles” of content are – that is, to what entities content properties are ascribed. Proposed vehicles include persons, subpersonal items (neural entities), or

something more than either of these, as in “vehicle” externalism.²⁰ I will assume the standard view that the vehicles are neural states, events or processes.

1.3. MENTAL ONTOLOGY

The problem of mental causation is essentially connected to the ontological problem of the mind-body relation. Mind-body interaction requires an explanation only on the assumption that there is a significant mental-physical distinction to be made. Currently, some form of non-reductive physicalism is usually presupposed in the mental causation debate. In this section I will discuss the ontological issues that form the background of the problem of mental causation and that of externalism in particular. These include the role of descriptions, the various forms of mental realism, and the standard ontological categories used to classify mental phenomena. Overall, I will adopt mainstream positions; the main exception is that I leave open whether token-token identity is true. This is because Pereboom’s account of causal powers suggested in chapter 4 involves denying token-token identity.

1.3.1. Essential Properties and Causal Powers

There is one conflation that needs to be clarified before proceeding. This is regarding the use of ‘essential’ as a synonym for ‘intrinsic’ and ‘causally efficacious’.

It is often taken for granted that only intrinsic properties can be essential to particulars.²¹ Some philosophers rule out of hand the possibility of essential extrinsic

²⁰ Baker (1995), Hornsby (1986) and Steward (1997) defend a persons view, while Hurley (op. cit.) floats the idea of vehicle externalism. Hornsby (1997: 168) complains that there’s no place for her in Fodor’s “Guide to Mental Representation”, since one can be a realist about the attitudes because one is a realist about persons, not because one is a realist about contentful states occurring in bodies.

properties, while for others the assumption is simply presupposed in the course of drawing an overall or intuitive contrast between intentional (extrinsic, inessential) and physical (intrinsic, essential) properties.²² This assimilation of ‘intrinsic’ and ‘essential’ becomes problematic when ‘essential’ is assimilated to ‘causal’. For example, Fodor (2000: 24) suggests that a natural way of interpreting the computational theory of mind – the idea that mental processes are syntactic – entails what he calls Principle (E):

(E) Only essential properties of a mental representation can determine its causal role in a mental life.

He adds that ‘syntax’ is essential because it is “context invariant”: “and, of course, the context dependent properties of representations (or of anything else) are not among their essential properties.”²³ In this passage, the assimilation of ‘intrinsic’ (in the sense of ‘context invariant’), ‘essential’ and ‘causal’ is transparent.²⁴

²¹ For some philosophers, ‘intrinsic’ *just is* ‘essential’ or ‘constitutive’, on the assumption that ‘intrinsic’ does *not* imply ‘non-relational’. (Writers on relational properties (e.g. Francescotti (1999), Pereboom (1991) typically hasten to distinguish between intrinsic/extrinsic and essential/not-essential.) But given the usual definitions and uses of ‘intrinsic’ in the literature I discuss, I think it will be far less confusing to maintain the link between ‘intrinsic’ and ‘non-relational’ and use ‘constitutive’ in such a way that both intrinsic and relational properties can be constitutive. See 1.5 below for illustration of this point.

²² Baker (2000: 30 fn. 10) explicitly notes that she “departs from the tradition in holding that not all essential properties are intrinsic.” Bennett (1988: 12) contrasts descriptions of events that involve their “intrinsic nature” with those that involve their relational properties, while Steward (op. cit.: 32-33) rules out relational properties from playing an ontological role. Finally, Stich (op.cit.: 151) assumes intentional properties aren’t essential in the following passage: “[The Syntactic Theory of Mind] theorist need say very little about the essential nature of the state tokens which are causally implicated in the production of behavior. Presumably they are physical states of the brain, *and thus* the properties which constitute their essential types are neurological properties, though an STM theorist who wishes to be coy or cautious need not even commit himself on this point. (Since it will simplify my exposition considerably, I will henceforth suppose that the states causally implicated in the production of behavior are neurological.)” (My italics.)

²³ Fodor (2000:26) elsewhere explicitly contrasts ‘not intrinsic’ with ‘context-dependent’. However, he does allow that some syntactic properties are relational yet ‘context independent’: ‘adds x number of words to a text’ is a syntactic, relational, essential and yet context-independent property of a sentence, while ‘makes a text globally odd’ (= have an odd number of words in total) is a context-dependent property of the sentence. So the fit between ‘intrinsic’, ‘essential’ and ‘non-relational’ is not precise. But the way in which it is not precise does not allow that semantic properties are essential; it just allows for some relational syntactic properties.

²⁴ Horgan (1997: 173) also appears to make this assumption when he asks rhetorically (in relation to Kim): “If mental properties are not even constitutive [essential] properties of events, one might well wonder, how can they be *causal* properties?”

A minor point (in this context) to make about this conflation of ‘essential’ and ‘intrinsic’ is that it is mistaken. Intrinsic properties need not be essential to the individuals that have them – for example, an object’s current mass or a mental state’s syntax – and extrinsic properties can be essential to the particulars that have them.²⁵ The important point is that the issue of which properties, if any, are essential to particulars is orthogonal to the debate over causal powers. The class of a particular’s essential properties (should it have any) need not be coextensive with that of its causal powers or its intrinsic properties. There is a legitimate link between ontology and causation made in the claim that we have no reason to posit the existence of something that has no causal powers. Kim (1993: 348) codifies this link as (Samuel) “Alexander’s dictum”: “To be real is to have causal powers.”²⁶ But this is a separate issue. Causal powers need not be essential properties, and whether causal powers must be intrinsic is precisely what is at issue. The terminological conflation mentioned above tends to obscure this point.

There is of course a clear role for a notion of essential properties in this context. Essential properties play a role in externalistic individuation of contents, since wide contents are individuated in relation to kind essences, or what is essential to a kind or type. Thus, the kind we refer to as ‘water’ is essentially H₂O. But this is distinct from the notion of an essential property of an individual, which is one without which the particular would not exist.²⁷ Individuals may have essential properties; having two hydrogen atoms seems to be an essential property of a token water molecule. But the questions of whether

²⁵ Humberstone (1996), Francescotti (1999) and others writing on the intrinsic/extrinsic distinction (discussed at length below) routinely distinguish between intrinsic and essential properties. See also footnote 20 above.

²⁶ As Armstrong (1978: 11) puts it, if a property is causally idle, we have no good reason to postulate it.

²⁷ See Baker (2000: 35-39), MacDonald (op. cit.), Rey (op. cit.: 28-34). Burge (1995) seems to be referring to essential properties when he writes of the “fundamental forms of individuation” of mental states.

individuals have essential properties or not, and of which properties are essential to them if they do, is not of concern here.

I do not mean to claim that contents are *not* essential to the particulars that have them. The contents of mental states are often thought to be essential to their identity; as Peacocke puts it, externally individuated contents are “a matter of the identity of the state itself, and not some way of thinking about it.”²⁸ However, it is not always clear whether these claims are about what is essential to a given content *type* or what is essential to a *token* of a given content type. What matters here is just that some mental representations do have intentional properties, and that we’d like to know whether these properties are causal powers, whether they are essential to the tokens that have them or not.

1.3.2. Mental Realisms

Intentional realism is a form of mental realism, which claims that mental phenomena exist. One of the strongest reasons we have for thinking something exists is that it causally interacts with other existents. So a defense of mental causation is also ipso facto a defense of mental realism. This motivates making clear the different sorts of mental realism that are supported by a defense of mental causal efficacy; more importantly, different mental realisms in turn support different kinds of causal efficacy.

Mental realism can take two basic forms: realism about mental particulars or realism about mental properties. Since the latter is a form of property realism and I want to maintain neutrality on that issue, I will set aside this possibility; nothing hangs on it here. My focus on tokens is also motivated by McGinn’s Non-Locality argument (see

²⁸ Peacocke (1992: 203). Fodor (1987: 11) also notes that propositional attitudes “have their contents essentially” (although he also holds that our canonical method for picking out contents, via that-clauses, also individuates them).

Chapter 3), which appears to locate the problem of intentional causal efficacy at the token level, and by Pereboom's account of mental causal powers (see Chapter 4).

There are three basic sorts of realism about mental particulars. The claim that mental particulars really exist does not distinguish among them. They differ as to the relation of mental particulars to physical particulars and the existence conditions of both relata. To a first approximation:

(a) An *ultra-robust realist* about mentality holds that mental particulars exist and that they are composed of non-physical stuff. This is a Cartesian dualist view. The mental particular's constitutive property must be a mental property, since presumably it has no physical properties. Physical particulars of course have their own existence conditions.

(b) A *robust realist* about mentality holds that mental particulars exist and are composed of basic physical stuff, but that a mental particular is not identical to the composite of basic physical stuff of which it is made. On this view, the existence conditions of a mental particular are determined by its mental property, not by its material constitution. Some philosophers, such as Steward and MacDonald, consider constitution accounts just weak versions of identity, but this is false if constitution is not identity.²⁹ This sort of realist can (but need not) affirm a token identity thesis between mental particulars and particulars at a non-basic physical level (such as neural states of some sort), but she will deny that these higher-level states are token-identical to a composite of basic physical states. Pereboom is an example of this position.

(c) A *weak realist* about mentality holds that mental particulars exist and are identical to particulars made of physical stuff. This realist affirms that (molecular)

²⁹ Rea (1995) is an overview of arguments on constitution and identity. Other important articles include Johnston (1992), Pereboom and Kornblith (1991) and Pereboom (2002). See also Baker (2000). I'll be discussing these views further in Chapter 4.

constitution is identity. On this view physical constitution determines the existence conditions of the mental particular; the particular's mental property is not constitutive of the particular. The particular counts as mental because it can be truly described using a mentalistic description, but this is not a matter of its existence; its mentality is skin-deep, as it were.³⁰ Almost all philosophers who hold token-token identity appear to be weak realists (although some are also eliminativists – given that there does not seem always to be a firm distinction between claiming A just is B and eliminating A).

In sum, if constitution and identity are distinguished, then ultra-robust realism claims that mental particulars have mental constitution and mentally-determined identity, robust realism claims that mental particulars have physical constitution and mentally-determined identity, and weak realism claims that mental particulars have physical constitution and physically-determined identity. In the latter two cases the mental particulars also have physical properties or descriptions, but the ontological (constitutive) role of the physical properties differs.

Both robust and weak mental realism are forms of non-reductive physicalism; they agree that mental and physical properties are not type-identical and that mental tokens are composed of physical stuff, but differ on whether basic physical constitution is identity. Although there are various forms of non-reductive physicalism (see 1.7.2), robust realists are non-standard (or perhaps radical) non-reductive physicalists in that they deny both type-type and at least some claims of token-token identity.

³⁰ Stich's comment in footnote 23 illustrates the entailment that is implicitly drawn from (a) "mental tokens are physical tokens" to (b) "the constitutive properties of mental tokens are physical". Unlike the weak realist, the robust realist thinks that (b) does not follow from (a). A significant motivation for this position is discussed in 1.3.1.

In what follows I will assume a non-Cartesian intentional realism: there really are beliefs, desires and other intentional particulars that are physical at least in their basic constitution.³¹ For now I will leave open the choice between robust and weak realism – that is, between denying or accepting some form of token-token identity – but, as noted, in chapter 4 the metaphysical framework for robust intentional powers put forward by Pereboom includes denying certain token identity claims.

Another form of mental realism is a realism about particular persons, not particular mental states; Hornsby and Baker support this view. However, since almost all of the mainstream discussion on my topic assumes either robust or (almost always) weak realism, I will set this view aside.

1.3.3. Descriptions and Identity Conditions

Contents are typically referred to as properties of mental particulars, even though talk of properties (or types) does not imply ontological commitment to properties.³² Talk of descriptions (as in “intentional description” or “mentalistic description”) is also common, but description-talk no more implies nominalism than property-talk implies realism. Descriptions can only truly characterize, classify or apply to the objects they describe in virtue of features the objects have. Nothing in description-talk implies that

³¹ Fodor (1990: 5) defines intentional realism (realism about the propositional attitudes) as the claim that there are mental states that are both causally efficacious relative to behavior (in ways that respect the generalizations of folk psychology) and semantically evaluable (have truth or satisfaction conditions). Since this definition leaves open how these states are related to physical states (i.e. whether token-token identity is true), it can be used in conjunction with any of the realisms I defined above. So if Fodor’s criteria are taken into account, by “intentional realism” I will mean the claim that there are causally efficacious, semantically evaluable states that are constituted in physical states.

³² That said, it is also traditional to think of contents as relations to propositions (i.e. abstract objects). As noted above, this view of content as a relation to an abstract entity plays a role in some of the arguments against externalistic causes discussed in Chapter 3.

there is no such a thing as sameness of features between distinct particulars or that sameness of feature does not require a realist explanation.

Moreover, the fact that a particular has numerous true descriptions in no way implies that all the descriptions are equal. They are on a par as linguistic entities, but not in terms of the features of the object they denote. The ontological situation is parallel to the causal situation in the extensionalist view of causation (see 1.4.1). If a particular can be described one way in a causal explanation but may be subsumed under a causal law under another, the particular can sustain equally good descriptions in mental or physical terms without this sort of equality implying equality in the causal role of the features those descriptions denote; one good description denotes a causally efficacious property, and the other good description does not.

The same lesson applies here. A particular can be described in one way in an explanation, but its existence conditions need not be revealed in that description. If constitution is not identity, then the complete physical description of an entity down to the last detail at any given time would not denote a feature essential to it. In short, the availability of a physical description entails nothing about the role, if any, that the physical feature corresponding to that description plays in determining the particular's identity. A robust realist in particular can agree that mental representations always have physical descriptions but deny that these play ontological (constitutive) roles.

1.3.4. Events, Processes and States

Assuming intentional realism, the question arises as to which ontological kind these mental entities belong. The relevant literature typically includes reference to mental

events, states and processes (but not mental substances), more or less interchangeably. Paradigm cases of each might be my sudden realization that I forgot to water the plants, my being anxious over the outcome of an exam, and my drawing a conclusion. While these ontological categories are distinct, there is no reason to think mental phenomena cannot be found in each of them.

Events and processes are similar in that they are closely, perhaps necessarily, connected to the notion of change. Different views of events include (a) Kim's view that events are property exemplifications by subjects at times; (b) Bennett's trope variant of Kim, in which events are instantiations of complex properties at space-time zones; (c) Steward's view, adopted from Lombard, of events as non-relational changes in objects; and (d) Davidson's view of events as particulars individuated by their causal relations. Since Kim's property-exemplification view is most familiar and widely held, I will adopt that view whenever it may be necessary to specify what is meant by "event".

A feature of Kim's account already briefly noted is his distinction between the properties that are constitutive of events and those that are exemplified by an event that, so to speak, already exists. Particular events *are* exemplifications of (constitutive) properties, but particular events also *have* (characterizing) properties.³³ In these terms, the distinction between robust and weak realism about mental particulars amounts to whether one thinks intentional properties are constitutive or characterizing properties of particular mental events.

³³ According to Kim (1993: 33-52), a difference of constitutive property entails a different event. As MacDonald (op. cit.:103) put it, Kim's constitutive properties "determine the very nature of the events which exemplify them in a way that characterizing properties do not, for they figure in the existence and identity conditions of events." Baker (2000: 38) suggests that essential properties should just be called existence conditions generally.

Processes are variously analyzed as (a) composites of events (the standard view); (b) complex entities that are composed of but not identical to events (the Dretskean view); or (c) entities that are both ontologically distinct from events and not composed of events (Steward's view).³⁴ Dretske distinguishes processes from mere sequences of events by claiming that processes have a product; their individuation conditions involve reference to purposes. Steward argues that processes and events spatiotemporally coincide, but that events are count-quantified (like horses) while processes are mass-quantified (like gold). On either (b) or (c), however, processes are not reducible to events. Although I will assume that processes are composed of events, I will leave open whether or not they are mereological sums of their component events.

States are often thought of as conditions rather than changes or occurrences. However, it is unclear whether this intuition is indicative of a distinct ontological category. Kim's account of events does not make a significant ontological distinction between states and events, since something's being red is an exemplification of a property just as much as something's diving into the ocean. Others, such as Steward and Robinson, argue that while the use of the linguistic frame "a state of ..." is natural in some cases (e.g. a state of anxiety), its indiscriminate use in philosophy of mind has created an appearance of ontological conformity that is ungrounded. So talk of mental states may just be an artifact of linguistic flexibility which either does not indicate a distinct ontological category from that of event or does not indicate an ontological category at all. The term "token state" may just refer to a particular event (as defined by Kim), ontologically speaking. However, since the classification of mental phenomena as

³⁴Kim (1993: 184) provides an example of (a) -- the standard view -- when he defines an internal process as "a causally connected or continuous series of internal events or states involving the same object or system of objects." See Dretske (1988: 33-44) for his view and Steward (1996: 94-97) for hers.

states is so deeply entrenched, I will continue to use the term “state” while leaving open whether this marks a distinct ontological category from that of “event”.

1.4. CAUSATION AND CAUSAL EXPLANATION

If token beliefs (or other intentional mental phenomena) are causes, then they stand in causal relations and can be cited to causally explain the occurrence of other events. In this section I will clarify the main issues involved in claiming that something is a cause and that it causally explains something else.

1.4.1. Causal Relata and the Role of Properties

As Kim notes, events, objects, event-types and other entities all have “a natural place” in Hume’s analysis of the causal relation; but conditions, states, phenomena, processes, facts, property-instances (tropes) and aspects have also been suggested as causal relata.³⁵ In short, our myriad linguistic means of referring to causes do not reserve that status to a single ontological category. So if mental phenomena, and propositional attitudes in particular, fall into one or another of the categories of possible causal relata, they satisfy at least this initial intuitive requirement for being a cause.

This appears to be the case, since mental phenomena either are typically classified ontologically as events, processes or states (or conditions). I will remain neutral about the choice of ontological category for mental causes, given that all three of the categories in which mental phenomena are typically classified -- “event”, “state” and “process” -- are among those listed as possible causal relata.

³⁵ See Kim (1993: 3-21) and his article in Sosa and Tooley (1993). Bennett (op. cit.) suggests tropes, while L.A. Paul (2000) proposes aspect causation. I do not assume all these kinds are deeply distinguishable.

In paradigm cases, singular causal claims involve two singular terms referring to distinct particulars or individuals linked by the relational predicate “cause” (or a variant, such as “bring about” or “produce”) or by a specific causal verb (such as “broke” or “spilled”, which may be analyzed as “cause + infinitive”). Such claims are usually analyzed as having the logical form $C(x,y)$, where “x” and “y” are singular terms referring to particular events (“The explosion caused the fire”). Some have argued that the causal relation is a three-place relation, with implicit reference to an alternative cause, but I will ignore this suggestion here.³⁶

The standard analysis also leaves open whether the singular term referring to the cause picks out (a) a pragmatically selected sufficient part of the “whole cause”, which is really the entire complex of events and conditions that bring about the effect (the Millian view); (b) an event that is itself the cause, but which is a pragmatically selected necessary part of a sufficient condition for the effect, or an INUS condition (Mackie’s view); or (c) the whole (Millian-style) cause, whose status as being necessary or sufficient depends on how it or the effect is described (Davidson’s view).³⁷ I will remain neutral among these choices. This means that while I adopt the standard Davidsonian-inspired terminology of causal relations as involving events with mental or physical descriptions (see below), this does not imply the Davidsonian view of the reference of the singular terms in causal statements. The issue of whether a singular term in a singular causal claim picks out the whole or a part of the cause is distinct from the issue of whether the singular term is a

³⁶ See Hitchcock (1993).

³⁷ Davidson (op. cit.: 149-162) presents his view and discusses Mill’s; Steward (op. cit.: 145) also discusses Mill. Mackie (1965) presents his view, also discussed in Bennett (op. cit.: 44). Horgan and Tye (1985) discuss Davidson.

complete or incomplete description of that cause. An INUS condition can be described incompletely, and a Millian-style “whole cause” can (in theory) be described completely.

An important feature of causal relations that is widely accepted is that, in any given causal relation, particular causes have their effects in virtue of only some of their properties.³⁸ A rock breaks a window because of its mass and velocity, not because it was one of five rocks in the thrower’s pocket. Reference to a property is usually shorthand for reference to a particular’s having that property: the rock’s velocity, not velocity in general, is a cause of the window’s breaking. However, if events are property-instantiations (Kim’s view), there seems to be no difference between saying the causal relata are particular events or property-instances. I will put this together by saying that mental properties are causally efficacious iff (token) mental events enter into causal relations in virtue of their mental properties. This leaves open whether the properties are constitutive or characterizing, so it is neutral between robust and weak realism.

The relatively fine-grained way of picking out causes by reference to particular features or properties of particulars is compatible with the widely accepted “extensional” view of singular causal relations associated with Davidson. On this view, assuming that singular causal statements relate particular events, the singular terms (names or descriptions) referring to events in a true singular causal statement can be replaced by co-referring names or descriptions without altering the truth value of the statement. If descriptions pick out events by referring to one or another of their properties, then the

³⁸ See McLaughlin’s (1995) discussion of Davidson, where he argues that this is compatible with Davidson’s extensional view of causation (see below). Block (1990: 139-40) nicely illustrates the view: “Some properties of a cause are relevant to the production of an effect, and some are not. ... I shall say that property P of event *c* is causally relevant to effect *e* and that *c* causes *e* in virtue of P more or less interchangeably.” Fodor (1989: 64) is another clear example in his discussion of an event’s “causally responsible properties” – i.e. some property or other in virtue of which it is a cause.

claim that an event referred to in a singular causal statement by one description may enter into a causal relation under another description is equivalent to the claim that an event picked out by reference to one of its properties can enter into a causal relation in virtue of another.³⁹

If type-type identity is false, the extensionality claim implies that there can be a gap between the way a causal relatum is described in a singular causal statement and the property in virtue of which the relatum stands in a causal relation. For example, if the infection in the meat caused the outbreak of food poisoning, and the infection is the proliferation of *E. coli* bacteria, then the proliferation of *E. coli* bacteria in the meat caused the outbreak of food poisoning. However, if the infection is also truly described as the subject of the first rumor Dan Rather heard on his way to work on July 10, 1999, the causal claim remains true under that description. However, its property of being the subject of the first rumor Dan Rather heard on July 10, 1999 is not the property in virtue of which it caused the outbreak. Similarly, an event that is picked out by a psychological description may interact causally in virtue of a physical property that is not identical to the property of the event in virtue of which it was picked out by the psychological description.

Although the existence of this gap is not specific to mentalistic descriptions, the distinction between descriptions (properties) that are or are not causally efficacious in a given causal statement (relation) plays a very important role in discussions of mental causation. For it abets the intuitions of semantic causal inefficacy that motivate many of the arguments examined in the following chapters. Similar intuitions play a role in the

³⁹I will generally follow Bennett's (op.cit.: 51) recommendation to rework Davidson's talk of descriptions into the material mode. See 1.3.3 for my motivation.

widespread view that Davidson's anomalous monism entails epiphenomenalism about mental properties.⁴⁰ If causation is a matter of nomologicality and there are no psychological or psychophysical laws, then it seems to follow that the properties that matter for causal relations cannot be mental, even if the event in question is picked out by a mental description. However, this difficulty is not due to the extensionality of the causal relation, but to the other features of Davidson's account. In short, mental causal efficacy is consistent with the extensionality of causation.

1.4.2. Regularity and Counterfactual Views of Causation

Davidson's nomological criterion of causality is perhaps the most well-known contemporary example of a regularity (or covering-law) account of causation, which is usually associated with Hume. On a regularity view, for two distinct events C and E, C causes E iff the occurrence of C, given the right laws or lawlike regularities and in conditions D, is sufficient for the occurrence of E. In Davidsonian terms, two events are causally related iff there are true descriptions of them that are lawfully related, even if these are not the descriptions that appear in the singular causal claim. The causal relation is derived from the existence of laws or lawlike regularities that subsume C and E under some true description or other. Roughly, lawlike generalizations are those that support counterfactuals, are confirmed by their instances and have predictive power.

Davidson also specifies that the laws under which particular causal relations fall must be strict or exceptionless. Other regularity accounts, notably in Fodor (1989), relax this criterion and allow that strict, exceptionless laws are not necessary for causation; *ceteris paribus* laws will do.

⁴⁰ Sosa was one of the first to make this criticism. See also McLaughlin (1989).

The main alternative analysis of the causal relation is the counterfactual view, originally associated with David Lewis and defended in various forms by Peacocke, Burge, Baker and Steward.⁴¹ In the version of this view in which the relata are events, two events are causally related iff the appropriate counterfactuals come out true. More specifically, for any two distinct events C and E, if C had not occurred, then E would not have occurred. There is causal dependence iff there is counterfactual dependence. The truth values of the counterfactual conditionals are usually determined (in technical terms) by looking at the closest possible worlds where the antecedents are true and seeing if the consequents are also true at those worlds.

A less well-known view of causal relations holds that they can be physicalistically reduced to (identified with) physical relations such as transfers of energy or momentum.⁴² Sosa and Tooley suggest that this analysis fails for reasons independent of the mental causation debate: there may be worlds where there are causal relations but where there is no physical substance, or it would be incompatible with the popular view that the temporal parts of continuants are causally related, since some of these relations (i.e. between the temporal parts of an electron) may not involve energy transfers. In any case, this view is a restatement of the claim that all causation is physical. It does not affect the externalism issue.

Although many defenders of mental causation opt for a counterfactual account of causal relations, I will remain neutral between regularity and counterfactual accounts. This is because the critical issue (discussed below) of the “genuineness” of intentional causation is orthogonal to the debate between regularity and counterfactual accounts of

⁴¹ See Lewis (1973), Peacocke (1993), Steward (op. cit.), Burge (1995) and Baker (1995b).

⁴² Sosa and Tooley (op.cit.: 3) briefly discuss the physicalist view suggested by David Fair and others.

causation. Fodor (1995) shows that a regularity view is compatible with a defense of genuine mental causes, while Schiffer (1991) shows that a counterfactual view is compatible with the claim that there are no genuine mental causes.⁴³ It may, however, be convenient to take one or another account for granted in a given instance. For example, what Fodor (1994) calls the implementation problem – the problem of correlating psychological laws with computational laws, given that the semantic and syntactic properties cited in the respective laws don't match up – is importantly related to the problem of externalism on the assumption that intentional causation involves the subsumption of intentionally-characterized events under intentional laws.

1.4.3. Causation and Causal Explanation

The extensional view of causation relies on a distinction between descriptions that can appear in singular causal claims and those that, in addition, denote causally efficacious properties of tokens. This is one way of expressing a certain relation between causation and causal explanation. As Strawson (1985) puts it, causation is “a natural relation which holds in the natural world between particular events or circumstances”, while explanation (causal or otherwise) is “an intellectual or rational or intensional relation [that] does not hold between things in the natural world, things to which we can assign places and times in nature.” One relation holds between natural entities whether or not we exist; the other is a conceptual relation between linguistic entities (or perhaps propositions) when we find that the one illuminates the other.

⁴³ In fact, Fodor (1990: 178-185, and fn.1) motivates intentional realism with the claim that there are intentional laws: since there are laws, there must be states that have the properties related in the laws. Schiffer argues that laws aren't needed to account for the explanatory role of mental properties (or any special-science properties): it's sufficient that they appear in true 'because'-statements to explain their “causal relevance”. The only laws we need are the strict laws of fundamental physics.

This distinction between causation and explanation, and therefore between causation and causal explanation, gives rise to two important issues in this context. The first issue is how causation and causal explanation are related. The most widely accepted view is that the former provides the ontological grounding for the latter; a true causal explanation counts as causal because “there is, behind it, an instance of causation.”⁴⁴ In the best of cases, the causal relation that grounds a causal explanation holds between events in virtue of those properties denoted by the predicates that play the appropriate roles in the explanation. However, as the extensional view of causation shows, the features in virtue of which a certain causal relation holds need not be mentioned even implicitly in the explanation. A causal explanation can be ontologically grounded in a causal relation even if it does not specify its ontological ground by referring to the property of the object that is causally efficacious in that relation.

The second issue is whether good causal explanations require laws.⁴⁵ The view that they do (which is not universally accepted) dovetails with the nomological account of causation. An explanation of why the window broke (because a rock hit it) cites one event to explain the occurrence of the second. These events stand in a causal relation in virtue of the fact that they can be subsumed under a law linking certain physical characteristics of the rock to certain structural characteristics of the glass. The same law can also be cited in a more complete explanation of the window’s breaking. So the causal law that subsumes two events also plays double duty as the law that is a necessary part of

⁴⁴ See Steward (op. cit: 169). Kim (1993: 254-260) endorses this view with his “causal realism” and “explanatory realism”. The first is the view that “there is an objective fact of the matter about the causal history of any given token event”, the second that “having a correct causal explanation of an event (as specified or described by an explanandum) requires that the token event described by the given explanans as the cause be, *in reality*, a cause of the event described by the explanandum.” (My italics.)

⁴⁵ Hempel (1948) began this tradition with his deductive-nomological model, although it is associated most closely with Davidson in the special case of mental causation.

the explanans. The counterfactual account of causation, which does not require law subsumption, fares better if good causal explanations involve the appropriate true counterfactuals but do not require laws.

Psychological causal explanation is a battleground for both issues. Since the success of such explanations is widely accepted (but see 1.7.1), debate has focused instead on the criteria for specifically causal explanations and whether psychological explanations meet these criteria. Fodor (1989) exemplifies the traditional view of causation as involving law-subsumption in the case of psychology and intentional properties: for an intentional property to be causally responsible is for it to be a property cited in a psychological law an instantiation of which nomologically necessitates an instantiation of the effect-property. A psychological causal explanation is an explanation grounded in a singular psychological causal relation as so defined. Burge (1995) and Baker (1995), however, urge that explanation comes first, such that an explanation is causal if we accept it as such. This position reverses the usual account in which causal explanations count as causal if they are grounded in causal relations. Baker, for example, explicitly rejects “the metaphysical picture [of materialism], which subordinates explanation to causation, where causation, in turn, is conceived as an ‘objective relation’ in nature.”⁴⁶

This position also implicitly rejects the demand for laws as essential to good causal explanations, for two reasons. First, the motivation for the demand for laws in explanation stems at least in part from the fact that the laws cited in explanation are the laws that subsume events in natural causal relations. By rejecting the idea that causal

⁴⁶ See Baker (1995: 93). In her terms, “causes are the sorts of things that are cited in explanations of events.”

explanation is causal because it is grounded in natural causal relations, this motivation for requiring laws in explanation disappears. Second, many accepted psychological causal explanations (like many explanations in general) do not cite laws.

Reductive physicalists have tended to maintain the traditional connection between causal explanations and their ontological grounds and to deny that intentional causes are real or robust or genuine. The two ways in which this has been done is to draw a distinction between (a) genuine vs. second-class explanations or (b) causally efficacious vs. causally relevant (or causal-explanatory) properties. In (a), the labels “genuine”, “serious” and “scientific” are reserved for those causal explanations that are grounded in relations between events in virtue of their physical properties. Intentional explanations may be good and useful, but they are not “genuinely” causal because the causally efficacious properties that play a part in the ontological ground of the explanation are not the intentional properties cited in the explanation.⁴⁷ In (b), it is granted that psychological causal explanations cite causally *relevant* (or causal-explanatory) properties, but the only causally *efficacious* properties (or “genuinely” or “robustly” causal properties) are those of physics (or those reducible to physics).⁴⁸ In the version in which causation relates

⁴⁷ As Pettit (1986: 19) explains the position: “Broad intentional states may be of social or semantic importance, but, failing explanatory indispensability, they will not be counted in any sober inventory of the psychological realm.” For criticisms of the “scientific” or “genuine” status of intentional psychology and its explanations, see Stich (1978: 590-592), (1983: 165-170, 139 fn.), (1991: 240), Fodor (1987: 39-42), (1991), Field (1986: 78-85) and Quine (1960: 219).

⁴⁸ Jackson and Pettit’s (1988) “programme explanation” view holds that causal explanations in a mentalistic (though not specifically externalistic) vocabulary depend on genuine causal relations between events under physical descriptions (or in virtue of their physical properties). The mentalistic descriptions are proleptic in that they point to a level of causal explanation and causation statable in a purely physicalistic vocabulary. Proleptic explanations can, however, be indispensable – irreplaceably useful. See also Millikan (1995) and Kim (1984). Steward (op. cit.) and Kim (1998) discuss the “proleptic” view. Similarly, McDowell (1985), following Davidson, argues that psychological explanations are “explanations in which things are made intelligible by being revealed to be, or to approximate to being, as they rationally ought to be”, as contrasted with e.g. Humean-style nomic subsumption. But he allows that the events in “ideal-involving” explanations can always in principle be given a Humean-style explanation.

facts, both mentalistic and physicalistic facts can play causal-explanatory roles, but the level of natural causal relations is restricted to physicalistic causes.

These issues will be touched on in later chapters. However, in what follows I will assume that if contents are causal powers, they are genuine or robust causal powers. I see no reason to fight for or accept any lesser grade of causal efficacy. (Thus, the claim that there is only one, first-rate, type of causal power will make no difference here.)

I also assume a view of intentional causation in which true intentional causal explanations are grounded in causal relations in which mental particulars play causal roles in virtue of their intentional properties – that is, a view similar to Fodor's. This is just a case of the standard view of the relation between causal explanation and causation, in which explanations are causal when they are ontologically grounded in natural causal relations. I have two reasons going this route.

First, the strongest arguments against the causal efficacy of wide intentional properties – the Non-Locality arguments of chapter 3 – focus specifically on the nature of relational properties. A position which claims that psychological causal explanations are causal because we accept them as such simply sidesteps these arguments; objections to the relational nature of contents are not even acknowledged. I think it is preferable to address the objections.

Second, any less robust account of intentional causation will be vulnerable to renewed objections questioning the reality of intentional causal relations, thus providing reason to deny intentional realism. At the very least, I think it is worthwhile to see if there is an account of intentional causation that does not require rejecting the standard view of

the relation between causation and causal explanation.⁴⁹ I think the account provided in Chapter 4 meets this criterion.

1.5. EXTRINSIC PROPERTIES

If externalism is true, then the semantic properties of mental representations are at least partly individuated in terms of the nature(s) of phenomena external to the thinker. This idea is often expressed by saying that wide contents are relational or extrinsic properties of mental states, as opposed to their internal or intrinsic properties. Chapters 2 and 3 will critically discuss arguments in which this characterizing feature of wide content is alleged to be the source, directly or indirectly, of the problem of externalism. However, these arguments rely on an intuitive understanding of the intrinsic/extrinsic distinction, one that is too vague for the central role it plays in them. For there are several distinct ways of drawing the intrinsic/extrinsic distinction, and hence several non-equivalent ways of expressing the claim that the contents of ordinary psychological states can't play causal roles because of their "extrinsic" (or "relational") nature.

In this section I'll present the main alternative ways that have been suggested for drawing an intrinsic/extrinsic distinction. I will defend a particular definition which I believe is clear, well-grounded in intuition and in the literature, minimally problematic,

⁴⁹ I also agree with Fodor (1989: 72-73) when he argues (against LePore and Loewer (1987)) that a counterfactual account does not satisfy our causal intuitions, which call for a cause to be nomologically sufficient for its effect. On LePore and Loewer's account, a mental property can be causally *responsible* for a given behavior, without being causally *sufficient*, as long as one of a disjunction of physical realizers is sufficient for the behavior. The counterfactual " $\sim P_m \& M_m \rightarrow B_b$ " can be true when M is nomologically sufficient for B (Fodor's view) or when M could have been realized by P* and P* is nomologically sufficient for B (L&L's view). In effect, L&L implicitly endorse a nomological sufficiency account, but allow that only physical properties satisfy it. Their view thus turns out to be yet another version of the "causally relevant but not really causal" position sketched in the text. The moral seems to be that we *do* demand nomological sufficiency (as Fodor claims), but that when we don't have it in the form of a strict law, we still have to account for the success and apparently causal nature of special science explanations, including those of psychology. So we can go with *ceteris paribus* laws with satisfied *ceteris paribus* clauses (Fodor) or we effectively relegate everything above physics to secondary status (L&L).

and sufficient to express the problem of externalism. I'll also clarify the relation of extrinsic properties to functional and mere Cambridge properties. One consequence of my choice of definition will be seen in chapter 3 in my discussion of McGinn's Non-Locality argument, which (I claim) relies essentially on a problematic understanding of the nature of extrinsic properties. The groundwork for that response is laid here.

1.5.1. The Intrinsic/Extrinsic Distinction

Given that almost all philosophical attention is directed at intrinsic properties, relational or extrinsic properties are often initially characterized negatively as those that are not intrinsic. But there is no one universally accepted definition of intrinsicness, and so no one for extrinsicness either. Perhaps there isn't just one. In any case, for each of the intuitive conceptions of intrinsicness that are sharpened in the competing definitions, there is a corresponding characterization of an extrinsic property. The question is which of these is most adequate in the context of the problem of externalism.

One adequacy condition is that it ought to be sufficient to express the problem of externalism without raising metaphysical controversies that are irrelevant in the present context. The definition should be neutral regarding the outcome of these further debates. A second adequacy condition is that the definition should fit the distinction between "syntactic" and semantic properties in philosophy of mind (such as in Block (1990) and Fodor (1994)). In that literature, the syntactic/semantic distinction is a special case of the intrinsic/extrinsic distinction. Fodor, for example, expresses this claim via the analogy of being dollar-looking (a.k.a. "syntax") and being a dollar (a.k.a. semantics):

Being a dollar is an extrinsic (causal/historical) property; whether a thing has it depends essentially upon its etiology. Being dollar-looking, by contrast, is a

matter of a thing's *internal* properties (it's a property of a thing's *appearance*, whatever, exactly, that may mean.) Being dollar-looking isn't even metaphysically sufficient for being a dollar since a thing's appearance may be so contrived as to obscure its etiology; it's here precisely that the forger inserts his wedge. (1994: 18-19)

The problem, Fodor goes on to argue, is that because semantics is external and "syntax" is internal, semantics and syntax need not be coinstantiated and so psychological laws can't be implemented in syntactic laws. This argument (as with many others) presupposes that the syntax/semantics distinction is a case of an intrinsic/extrinsic distinction, where "syntax" is intended to indicate neural structures at some level of abstraction.⁵⁰

Before going further, I should clarify an ambiguity in the term "extrinsic" (or "intrinsic") that has been called the global vs. local distinction.⁵¹ A global notion of "intrinsic" or "extrinsic" is a claim about kinds – whether a property or kind is intrinsic or extrinsic, as a matter of its (metaphysical) nature. A local notion is a claim about particulars – whether a particular has a property in virtue of its intrinsic nature or not. Most philosophical attention has focused on defining global notions, although the two notions can be interdefined: for example, we can define a globally intrinsic property as one which is locally intrinsic to all its possessors. But the notions are distinct: a property P can be locally intrinsic to x but not to y, in which case P is not globally intrinsic. For

⁵⁰ That Fodor has neural (or at least physical) properties in mind seems implicit in his following expression of the problem of how "syntax" can implement semantics:

How could a process which, like computation, merely *transforms one symbol into another* guarantee the causal relations between symbols and the world upon which ... the meanings of symbols depend? I can ... transform the symbol 'dog' any way I like: I can write the word backwards, or cut off the 'd' or replace it by the word 'cat'. But it is self-evident that no operation of *those* kinds will suffice to connect tokens of 'dog' causally with tokens of instantiated doghood. (1994: 13)

The representative "syntactic" transformations that Fodor provides are all physical – they involve changes in what Devitt (1996) calls a symbol's "brute-physical" properties. They are not (e.g.) transformations from one syntactic category (being a noun) to another (being a copula). So when Fodor says semantic properties aren't coinstantiated with syntactic properties, the "syntax" he has in mind is physical (e.g. neural).

⁵¹ The label is Humberstone's (1996). The distinction is also cited by Francescotti (1999). Both attribute it to Dunn (1990).

example, if P is the (disjunctive) property of being either made of tin or adjacent to something which is made of tin, then P might be locally intrinsic to Henry (if Henry is made of tin) and locally extrinsic to George (if George is wooden but next to Henry), and P itself would not be globally intrinsic (if an intrinsic property is one that is intrinsic to all its possessors). Global intrinsicness (or extrinsicness) does not follow from local intrinsicness, nor vice versa.⁵²

The global/local distinction plays a role in Ch. 4 in my response to an objection to the claim by some theoretical chemists that molecular structure is extrinsic. (This claim in turn plays an important part in my defense of wide semantic causal powers.) I respond that when we wonder whether certain properties *in science* are intrinsic or extrinsic, it still makes sense to talk of them as being intrinsic or extrinsic in the local sense – for example, in just the physically possible worlds – even while we leave open the question of whether they are globally intrinsic or extrinsic. Also, arguments against extrinsic causal powers that depend on claims about science (see especially Fodor’s Non-Locality argument in Ch. 3) only need a local notion to get going. As long as wide contents are extrinsic in the physically possible worlds, and causation is “local” in those worlds, Fodor can raise his objection. Of course, by the same token, if I can show that wide contents can be causes in our world, this would suffice as a response.

Thus, the claim that contents are extrinsic properties is at least the claim that they are extrinsic properties for at least some of their possessors, and the problem of externalism is at least the problem of whether a content property that x has extrinsically makes any difference to x’s causal powers. It will also do no harm to assume in addition

⁵² It may be argued that such disjunctive properties can be reduced to each disjunct, and each disjunct is intrinsic or extrinsic by nature. But there are other examples: being the same height as Arnold is a property that Arnold will have intrinsically and other people will have in relation to Arnold.

that contents are globally extrinsic, as writers on this topic appear to assume anyway. The question of externalism can then be subdivided into two – first, what distinguishes x’s intrinsic and extrinsic properties, and second, whether this difference maps onto a difference in x’s causal powers. This section addresses the first question.⁵³

Relational/non-relational. One way of drawing an intrinsic/extrinsic distinction is in terms of a non-relational/relational distinction. Intuitively, relational properties are those which somehow involve or consist in a relation.⁵⁴ For example, x’s being a brother might be said to consist in x’s having a certain relation to at least one sibling. A general definition of this notion might be as follows:

(R) P is a relational property iff $\forall x(Px \rightarrow \exists y(Rxy))$,

where “consists in” is interpreted as strict implication. However, this is a weak condition, since if R is the identity relation, the definition subsumes any property whatsoever. A stronger definition would eliminate this problem:

(R*) P is a relational property iff $\forall x (Px \rightarrow \exists y(y \neq x \ \& \ Rxy))$.

This is Bennett’s definition.⁵⁵ However, the inclusion of non-identity in this definition moves it partly towards capturing a very different idea than that of involving a relation: the idea that x and y are mereologically disjoint. (I’ll argue below that the most precise definitions focus on this idea, setting relationality aside.) I say “partly” because any property would be relational under (R*) if x has parts, for it would be satisfied if y is a

⁵³ In this section I rely somewhat on Humberstone’s (1996) extensive discussion of the various definitions that might qualify as an adequate way of drawing an intrinsic/extrinsic distinction. I omit his discussion (and rejection) of a definition based on relevance logic, developed by Dunn, in which a “relevant” property is one that at least depends on the nature of an individual. (For the record, Humberstone also notes that being intrinsic is neither necessary nor sufficient for being essential. See my discussion in 1.3.1 above.)

⁵⁴ There is also a distinction drawn between “impure” relational properties (those that involve relations to particular entities, places or times (e.g. voting for Arnold)) and “pure” relational properties (those that are purely qualitative and do not mention particulars (e.g. voting for someone)). The pure/impure distinction will not matter in this context.

⁵⁵ Bennett (1971: 253).

part of x and R is a constituency relation. Moreover, if x and y are both parts of the same whole, then any property will be relational given any R between x and y . In particular, syntactic and semantic properties would equally satisfy (R^*) . In short, the relationality definitions seem to miss the target concept being defined. If the distinctness of y is what's essential to extrinsicness, we might do better to focus on that (as Kim does, below).

Francescotti's (1999) develops a relationality definition in which extrinsic properties are a special case of relational properties; they are d-relational (d for "distinct", as in y 's being distinct from x):

(DR) F is a d-relational property of $x \leftrightarrow (\exists R)(\exists C)(Fx \text{ consists in } (\exists y)(y \in C \ \& \ Rxy \ \& \ x \neq y))$.⁵⁶

That is, F is a (pure) d-relational property iff: there is a relation R , and a class of items C , such that (i) x 's having F consists in there being some member of C to which x bears R and (ii) at least one member of C to which x bears R is distinct from x . However, (DR) also will not suffice to draw the syntactic/semantic distinction; F can be a syntactic property if x and y are distinct neural states of the same person.

However, even (DR) suffers from a second major difficulty faced by any of the relationality definitions: they engender metaphysical controversies that are unnecessary in this context. There is simply no widely accepted interpretation of the intuitive idea that an extrinsic property "consists in" or "involves" a relation.⁵⁷ It cannot be assumed that we

⁵⁶ An impure d-relational property is one such that the relation involves a particular distinct individual: (a^*) There is a relation R , and an item y , such that (i) x 's having F consists in x 's bearing R to y and (ii) y is distinct from x . Symbolically: $(\exists R)(\exists y)(Fx \text{ consists in } Rxy \ \& \ y \neq x)$. *De re* contents (see 1.2.3) are one kind of impure d-relational property. Pure d-relational properties are defined as above. Externalistically individuated contents that involve instances of constitutive properties – natural kinds, for example – would be pure d-relational properties. Francescotti classifies narrow contents as intrinsic properties of the individual who has thoughts with those contents. That is, he *implicitly* interprets his definition of d-relational property as quantifying over *persons*, not token mental states. Presumably he would also classify syntax as intrinsic for the same reason. But this interpretation is not an explicit part of the definition.

⁵⁷ Francescotti (1999) asserts but does not defend the "identity" interpretation.

can simply identify relational properties with relations. The identity interpretation is held by Francescotti (without defense), and is assumed in the following definition from Pereboom (1991: 343):

F is an extrinsic [relational] property of x just in case F is a relation x has to a thing wholly distinct from x.

This type of definition faces a serious conceptual difficulty. If F is an n-place ($n \geq 2$) relation, it is not the sort of thing that x can instantiate; it is something that only x and at least one other y can instantiate. But if F is a property of x, then it *is* something x instantiates. So the definition fails to explain how F could be both a property and a relation (or, if properties are 1-place relations, how relational properties could be both 1- and 2-place relations).

Yet if “consists in” is not identity, we have several choices. In (R) and (R*), “consists in” is interpreted as implication (“ $\rightarrow$ ”), although in this case if R is identity all properties turn out to be extrinsic. A stronger relation is also difficult to defend. If “consists in” is necessary equivalence or coextensiveness (“ $\leftrightarrow$ ”), this eliminates any asymmetry between relational properties on the one hand and relations and intrinsic properties on the other. For example, while the (intuitively relational) property of being made of the same material as something made of tin is necessarily coextensive with the (intuitively intrinsic) property of being made of tin, we would not want to say that being made of tin consists in being made of the same material as something made of tin.⁵⁸

Another metaphysical issue is that some relational properties can be complex, or, as Lewis puts it, “more or less extrinsic”: “being a brother has more of an admixture of

⁵⁸ I borrow this example from Humberstone, who also argues that without restrictions on what can count as a “real” relation, intuitively intrinsic properties (“is made of tin”) can be substituted for “P” above if we construct complex relations (“is made of the same material as y & y is made of tin”) with which they are necessarily coextensive.

intrinsic structure than being a sibling does, yet both are extrinsic.”⁵⁹ This complexity is common: being a pencil involves not just functional relations to human intentions, but also containing graphite and having an appropriate shape. Mental properties may also be complex, as shown by the possibility of what Rey (1997: 193-4) calls “anchored” functionalisms – proposed versions of functionalism in which mental properties are defined not just in terms of causal relations between inputs, outputs and other mental states but also by physiological constraints. The identity interpretation of “consists in” in particular seems to preclude these possibilities by definition.

Duplication. Lewis has developed an intrinsic/extrinsic distinction based on the intuition that an intrinsic property is purely qualitative – one that is shared by all duplicates, which are (on Lewis’s view) worldbound individuals that are exactly qualitatively alike. Those not shared by all duplicates are extrinsic. Lewis also defines the duplication relation in terms of intrinsic properties. He offers a pair of interdefined notions as a starting point for further elaboration:

(I) P is an intrinsic property $\leftrightarrow \forall x \forall y (D(x,y) \rightarrow (Px \leftrightarrow Py))$

(D) x and y are duplicates $\leftrightarrow \forall P (I(P) \rightarrow (Px \leftrightarrow Py))$ ⁶⁰

This sort of distinction seems to be the sort that (e.g.) Stich (1983) has in mind when he claims that replicas that differ in the contents of their beliefs will have the same causal

⁵⁹ Lewis (1983). This complexity is consistent with suggestions by Humberstone (op. cit. 224-5) and Armstrong (1978: 79) that relations are constituents of relational properties, although both also argue that relational properties are either reducible to relations and non-relational properties (Armstrong) or they correspond to one-place predicates defined in terms of two-place predicates, where the two-place predicates reveal the constituent conceptual structure of the one-place predicate and where predicates differ if their constituent concepts differ (Humberstone). Perhaps; but it would be preferable to avoid these issues altogether with a definition that is weaker but adequate for current purposes.

⁶⁰ I simplify Humberstone’s “Intr” to “I” and his “Dupl” to “D”.

powers.⁶¹ Also, Twin Earth scenarios rely on the notion of a duplicate or replica to argue for externalism. Oscar and Twin-Oscar are duplicates who presumably are identical in syntactic properties but differ in semantic properties.

I think there are two reasons – one technical, one conceptual – why an intrinsic/extrinsic distinction based on Lewis’s duplication relation is inadequate. First, duplication is understood by Lewis as a relation between numerically distinct, mereologically disjoint (worldbound) objects. This built-in modal realism enables the duplication-based definition to avoid the problem that arose when R was identity in the relationality definitions. An adequate intrinsic/extrinsic distinction should not depend on it. For example, in Burge’s thought experiments involving counterfactuals, whether in our counterfactual judgements the Patient is compared with himself or with a counterpart of himself will depend on whether we think there are transworld individuals or worldbound individuals. An externalist should be able to express the idea that the Patient has all the same intrinsic properties actually and counterfactually without having to commit to either view. So a duplication-based intrinsic/extrinsic distinction should not build modal realism (worldbound individuals) into the notion of duplication if it is to be used to define “externalistic” contents. However, if D can be (transworld) identity, all properties turn out to be intrinsic by the above definition, including syntax and semantics. Nor is it clear that we can fix matters by adding the clause “& $y \neq x$ ”, since this seems to force us to deny that we’re just imagining the Patient – that very same guy – in a counterfactual situation. We should prefer a definition that avoids these metaphysical headaches.

⁶¹ Humberstone clarifies that replicas and duplicates are not the same: replicas are physical-duplicate world-mates, i.e. within one world. So they are a special case of duplicates. But the definition is based on the duplication relation, not the replica relation.

There is a deeper problem that seems likely to affect any “purely qualitative” intrinsic/extrinsic distinction, of which the above is just a prime example. Suppose we hold that duplicates of x have the same thought contents; that is, Oscar in @ and Huascar in W_2 , where there is water, count as duplicates, but Oscar and Twin-Oscar do not. Then content properties are intrinsic by the above definitions, along with syntactic properties. That is, we determine which properties are intrinsic by selecting which entities count as duplicates, but without any explicit restrictions on our selection criteria for duplicates, any property can be intrinsic. Of course, Lewis explicitly invites us to use his “tight little circle of interdefinables” as just a starting point. The problem is that the above definitions aren’t doing any work in drawing the intrinsic/extrinsic distinction. What’s doing the work is either our implicit selection criteria for what counts as a duplicate or our implicit idea of what counts as an intrinsic as opposed to an extrinsic property. If we made either of these implicit ideas explicit – particularly in the latter case – we’d already have all the definition we need, and (I) and (D) would be superfluous.⁶²

Internal/external. Lewis himself tries breaking into the circle (or, I think, looks for a better definition) by examining a suggestion of Kim’s that starts from the intuition that an intrinsic property is one that its possessor has in and of itself. (He eventually adopts Kim’s definition, with some minor modifications.⁶³) This intuition behind the

⁶² Francescotti, following Dunn, raises the problem that since the duplication relation is transitive, C ’s being a duplicate of A turns out to be an intrinsic property of C . So duplication seems insufficient for intrinsicness.

⁶³ See Langton and Lewis (1998).

internal/external definitions is often not distinguished from those behind the relationality definitions, although the resulting definitions are not equivalent.⁶⁴

In Kim's formulation, an internal property is one that does not imply the existence of anything other than its possessor:

(T) P is internal $\leftrightarrow$ Possibly some object x has P although no contingent object wholly distinct from x exists.⁶⁵

In turn, by this global definition, an extrinsic property would be one such that, for any object that has it, there must exist at least one numerically distinct, mereologically disjoint object other than the possessor of the property. Roughly,

(E) P is external $\leftrightarrow \forall x (Px \rightarrow \Box \exists y (y \neq x \ \& \ y \in x))$ ⁶⁶,

where “ $\in$ ” is non-constituency. Unlike the relationality definitions, (E) emphasizes the the implied existence of a distinct object y , not the relation that x has to y .⁶⁷ Also (as Francescotti also notes) the object y typically has a particular nature; for example, x 's being a water thought entails the existence of at least one sample of water in x 's environment. So we can add a clause that there is some class of which y is an instance:

(E+) P is an external property $\leftrightarrow \forall x (Px \rightarrow \Box \exists y \exists Q (y \neq x \ \& \ y \in Q \ \& \ y \in x))$.

⁶⁴ For example, Lange (2002: 18) describes the intrinsic properties of a thing “roughly speaking” as those which “depend only on itself, not on its relations to other things”, while “a typical non-intrinsic (“extrinsic”) property is the property of standing in a certain relation to something else.”

⁶⁵ Humberstone (op. cit., 229), citing Lewis (1983: 198), who in turn presents a simplified version of Kim (1982/1993: 184). Kim begins with the idea inspired by Chisholm of an external property F as one being “rooted outside” the objects that have it: df. Necessarily any object x has F only if some contingent object wholly distinct from x exists. The local version (from Francescotti) is: F is an intrinsic property of $x =$ dr. x has F , and it is possible both that x has F and that no contingent object y wholly distinct from x exists.

⁶⁶ Lewis's version (stated in Pereboom (1991) is: F is an extrinsic property of x iff x 's having F *logically* entails or rules out there being a contingent thing wholly distinct from x . The “ruling out” part is to allow for negative extrinsic properties, on the assumption that extrinsic properties are closed under negation. If “logically entails” is intended to mean “entails in all possible worlds” it would be stronger than the intended reading of (E).

⁶⁷ Following Francescotti's symbolism for converting a global notion to a local one, the local definition would be: (E_L) P is an external property of $x \leftrightarrow Px, \ \& \ \Box \exists y (y \neq x \ \& \ y \in x)$.

Briefly, P is an extrinsic iff Px entails the existence of at least one distinct Qy. A local version of (E*) would state that P is an external property of x iff if x has P in W, then there must exist a Qy in W.⁶⁸

(E+) should actually be generalized to state that x's having P implies that x's world has a certain character, not that it contains a particular object. This is because in some important cases of extrinsic properties the characteristic of x's world on which x's having P depends is not the existence of another object in that world. It may be that x's world has n dimensions (in the case of chiral molecules, discussed in Chapter 2) or that it contains a vacuum electromagnetic field (in the case of molecular structure, discussed in Chapter 4). A general definition might go as follows:

$$(E^*) \text{ P is an external property } \leftrightarrow \forall x \forall W (Px \text{ in } W \rightarrow \Box \exists Q(QW))$$

In this definition, Q denotes some feature of x's world, such as "has n dimensions" or "is a y-containing world". If Q is the property of being a y-containing world, then we can analyze Q along the lines of the consequent of (E+) to get the special case definition.

There is still one difficulty: the externality definitions – like the relationality definitions – do not distinguish between y's being external to neural state x (e.g. if y is another neural state) and y's being external to the organism of which x is a neural state. However, this problem might be addressed with a clause specifying that x and y cannot be parts of the same whole.

Modulo this amendment, I propose to accept (E*) as a general definition of a (global) extrinsic property, and (E+) as a special case of (E*). My reasons are twofold.

⁶⁸ Also, entailment or implication is a relation between sentences or propositions, not individuals or property instances. (This goes for any of the definitions.) So strictly speaking, a true extrinsic description about a particular is one that entails a description of the world that the particular is in, while a true intrinsic description of an entity does not entail any true description of the entity's world.

First, the definitions satisfy the adequacy conditions set forth above (assuming the addition of the clause specifying that *x* and *y* are not parts of the same whole). They clearly capture the sort of state that is ruled out by Putnam's (1975) informal definition of methodological solipsism, the assumption that no psychological state presupposes the existence of any individual other than the subject to whom that state is ascribed. Wide contents are paradigms of non-solipstic psychological states. The definitions avoid the unnecessary metaphysical issues that plagued the relationality definitions. There is no troublesome "consists in" locution to deal with; these definitions are not attempting to reduce extrinsic properties to anything else. The definitions do not rule out complex extrinsic properties; they only say that those which are complex will be classified as extrinsic. They are also sufficient to raise in intuitive form the most serious problems raised for the causal efficacy of content, discussed in Ch. 3: if the external objects whose existence is implied by *x*'s having *P* differed (actually or counterfactually), would *x*'s causal powers differ? McGinn states the problem in just this way:

Surely, if the external relata of these relations were not to exist, this would not necessarily alter the causal power of states of the subject to bring about the same bodily movements. (1989:134)

In other words, why should the existence of a distinct contingent entity entailed by *x*'s having *F* make any difference to *x*'s causal powers? Why should the characteristics of *x*'s environment matter to the causal powers *x* has (and not merely their exercise)?

Second, the basic intuition expressed by these definitions is that an extrinsic property is one that can only be instantiated in a particular kind of world. Interestingly, this idea is *not* the complement of Kim's original motivating intuition of an intrinsic property as one an object has in and of itself. In fact, *neither* of (E), (E+) or (E*) is the

complement of Kim's original intuition. The definitions express instead the complement of the intuition that an intrinsic property is context-invariant or context-independent. An extrinsic property of an individual is one that entails something about the kind of environment in which that individual exists, while an intrinsic property of an individual has no such implications about the environment of that individual.

This intuitive conception of extrinsicness has much to commend it. It is close to the view in physics in which a property of a system (such as an organism) is a constraint or limitation on the possible states (behavior or development) of the system.⁶⁹ On this conception, an extrinsic property is a constraint on the individual's possible states that is at least in part due to the character of the individual's environment, while an intrinsic property is a constraint that is independent of the environment. For example, a neural state's being (partly) made of carbon is a constraint on the state – on its capacities – due to the characteristics of carbon, whereas its being a water-thought is a constraint due in part to the existence of water in its environment.

This conception also reveals an important similarity between extrinsic properties and *ceteris paribus* clauses. Both advert to scope limitations, the former in terms of the kinds of worlds in which a property can be truly ascribed to phenomena and the latter in terms of the kinds of worlds in which a law is true of phenomena. I return to this idea in Ch. 4 in my explanation of the animosity (among many philosophers, at least) to the idea of extrinsic causal powers.

Most importantly, this conception, and the resulting definitions, seem right on target in terms of capturing what is essential to extrinsicness. The relationality definitions conceive of extrinsicness in terms of the relations in which *x* stands. If the relation *R*

⁶⁹ See for example Ross (1999) and Lange (2002)).

differs so too does the relational property. The externality definitions conceive of extrinsicness primarily in terms of the properties instantiated in, or characteristics of, x's environment. These definitions include one relation: the entailment between x's having (or being) P and x's world's having (or being) Q. But the nature of this relation does not determine the extrinsicness of P. The nature of the environment does. This seems right. For when we say that a property of an object is extrinsic, we aren't saying anything specific about the relation between the individual and its world. We're saying something specific about the world – namely, that it has a certain character. When we say specific things about the relation, we are drawing distinctions among properties that have already been classified as extrinsic. The only relation *all* extrinsic properties *need* to have is an entailment from x's having P to x's world's being Q.

This difference in emphasis matters for several reasons. For one thing, we are not tempted to make a mystery of the fact that an extrinsic property is, after all, a property. The relationality conception makes this temptation almost overwhelming. The most obvious expression of this is when extrinsic properties are identified with the relations they involve. More generally, the idea that extrinsic properties involve relations – as opposed to the idea that they imply facts about the possessor's context – seems to lie behind at least some of the intuitive resistance to their causal efficacy. McGinn's Non-Locality argument (in Ch.3) is, I think, a case in point.

For another, we can express generalizations about extrinsic properties that otherwise might be missed. For example, correspondence theorists and deflationists can agree that a water-thought is extrinsic. On the relationality view, since water-thoughts stand in different relations for each theorist, they are distinct extrinsic properties. We

have no way to show what both views share – namely, the simple claim that water-thoughts are extrinsic. On the externality definitions, correspondence and deflationary theorists agree that x's having a water-thought entails the existence of water in x's world – that is, they agree on the extrinsicness of water-thoughts – and they disagree on the nature of the relation between water-thoughts and the world. Again, some theorists may think that wide content can be naturalized, others not. They agree on the extrinsicness of semantic properties, but disagree on how they are to be explained.

In the rest of this dissertation, I will continue to use the terms “externalistic”, “extrinsic” and “relational” in the usual ways they are used and in the way that I have used them until now – that is, as synonyms. However, whatever others may mean by those terms (and it is not always clear), I will mean (E*) or its special case (E+).

1.5.2. Mere Cambridge and Functional Properties

Mere Cambridge properties. The notion of a mere Cambridge property plays a role in discussions of my topic as a way of assimilating wide contents to a class of properties which seem clearly non-causal. As introduced by Geach (1969), if a change occurs to x when a predicate formerly true (false) of x comes to be false (true) of x, then *mere Cambridge changes* are changes that are not genuine or real.⁷⁰ Xanthippe's becoming a widow upon Socrates's death is an oft-cited example of a mere Cambridge change (in Xanthippe). Mere Cambridge *properties* are those acquired or lost by a particular through no material change in the particular. Being a widow is a mere

⁷⁰ That is, all changes are Cambridge changes (i.e. defined in terms of predicate satisfaction, or, in material mode, property gain or loss), but some of these are “mere”. See Kim (1993: 29) for discussion.

Cambridge property of Xanthippe since she acquires it through no material change in herself.

Genuine changes are often associated with those that involve the gain or loss of intrinsic properties. However, while some extrinsic properties are mere Cambridge properties, not all are. If *x* is taller than *y*, “is taller than *y*” may be a mere Cambridge property of *x* if *x* acquires it due to *y*’s growth, but it could be acquired by *x* due to *x*’s growth. So while “being taller than *y*” is an extrinsic property of *x*, it doesn’t follow that it is a mere Cambridge property of *x*.

Twin Earth cases may seem to suggest that wide contents are not just extrinsic, but are also mere Cambridge properties. One of the more celebrated rhetorical arguments against the causal powers of wide contents – Fodor’s (1987) example of H- and T-particles – makes the most of this appearance. Block (1986: 625) makes a similar (though non-rhetorical) claim when he says that differences in thoughts due to relational differences “do not cause any differences in proteins, synaptic connections, or any other physiochemical feature of the body, as do differences in genes and surface stimulations.”

I have no idea how Block could know this *a priori*. Externalism does not entail that external relations don’t have internal effects. It only entails that different external relations can have the same internal effects. The same internal effect *may* be the same *null* internal effect, but the nullness of the effect isn’t entailed by the extrinsicness of a property; the same internal effect might also be the same *non-null* internal effect. In the widow case there is presumably the same *null* effect on Xanthippe’s internal features when she gains or loses this property, but for all we know there may be the same *non-null* effect on Oscar’s and Twin-Oscar’s internal features when they gain or lose their content

properties. So it is a mistake to automatically assimilate content to the class of mere Cambridge properties.

There is also the possibility, cited above, that externalistic contents are a sort of “anchored” functionalist property, in which having a particular extrinsic content also involves certain physiological constraints. The plasticity of the brain – the way neural circuitry is formed and reformed in the course of experience – would seem to argue for non-null effects of environmental factors on internal features. This is all compatible with externalism. So rhetorical devices to the contrary, there is no reason to think contents are mere Cambridge properties.

Functionalism. The dominant view of mind-body relations which allows for realism about the mental is functionalism, or “metaphysical functionalism”, to use Block’s (1980) phrase. This view, in its purest form, holds that mental properties in general are functional properties – that a given entity exemplifies a mental property, or is truly describable as mental, iff it satisfies a certain causal role. A causal role is defined in terms of causal relations between inputs, outputs and other mental states. Mental properties are identified with this causal role; and mental particulars count as mental in virtue of their playing this role. This leaves open how the identity of the particulars is determined – that is, whether the role-property is just a way of characterizing the particular (“the President”) or whether it is constitutive of the particular (“the Supreme Court”). The “anchored” functionalisms contemplated in Rey (1997) are variations on this basic theme (what he calls “pure I/O”, or pure input/output, functionalism).

The main complaint about functionalism is that it does not seem to capture what is essential to mental states with phenomenal content (such as being painful or seeing

green). It is conceivable that a state might satisfy a certain causal role and yet not have the phenomenal character that intuitively is considered characteristic of that type of mental state. However, there are theories of phenomenal content (such as Dretske (1995)) and consciousness (such as Rosenthal (1991b)) that try to demystify the “what it’s like” intuitions that drive this complaint. In any case, functionalism in some form is generally accepted for intentional mental states, and I will assume it here.

“Relational” is not the same as “functional” because a causal role may be defined internally, such that an individual can satisfy the causal role in a way that does not imply the existence of an external individual. For example, narrow contents may be defined as causal roles in an individual’s mental economy, but since the inputs and outputs in the narrow causal role are internal, having a given narrow content in this functionalist sense does not entail the existence of any external individuals. So functional properties need not be relational in the sense of relational I am using.

Functionalism is thus neutral between content externalism and content individualism, since whether an intentional property is externalist or not will depend on whether “inputs” and “outputs” of the causal role are wide or narrow. If the inputs and outputs do not supervene on the agent, then neither will the causal role. A wide functional role theorist will hold that the inputs include perceptual experiences (which are externally individuated if Burge (1986) is correct) and the outputs include intentional actions (which are externally individuated if Millikan (1993) and Dretske (1988) are correct).

Functionalism is also neutral between robust and weak realism at the token level. The claim that a causal-functional role is satisfied by a physical property is ambiguous between a reading in which the functional property is a constitutive (essential) property

of mental tokens or one in which it is merely characterizes mental tokens. For example, saying, as Lewis does, that “the concept of pain ... is the concept of a state that occupies a certain causal role” is ambiguous between these readings. The two readings correspond to what Block (1980) distinguishes as “functional specification” functionalists and “functional state identity” functionalists. According to Block, the former claim that (e.g.) pain is functionally *specified* state, the latter that pain is a *functional* state. In constitutive vs. characterizing terms, functional state identity theorists hold that the constitutive (essential) property of a token is its functional property, not a physical property it may have. This is a robust realist position at the token level. But to functional specification theorists, a token whose physical property is constitutive can also be truly characterized by a functional property. This is a weak realist position at the token level.

Lastly, functional properties are often considered second-order properties in that their definitions quantify over first-order properties, as in Block (1990). On this view, a mental property is the property of being a certain causal role satisfied by other properties, themselves presumably physical in some broad sense; the mental property is the property of having a certain complex physical property. Since it is plausible (and perhaps actual) that the physical properties satisfying the same causal role can differ, functionalism is often (but not invariably) considered a form of non-reductive physicalism about mental phenomena. For if mental properties are multiply realized by distinct physical properties, they cannot be identified with one physical property and type-type identity is false.

Standard reductionist responses to this claim are that functional properties reduce to a (possibly open-ended) disjunction of physical realizer-properties or that a functional property can be identified with the physical realizer-property that satisfies the causal role

in a given instance.⁷¹ However, Pereboom (2002), whose “robust nonreductive physicalism” provides the basic framework for the positive account of semantic causal powers offered in chapter 4, argues that the realizers themselves can be mental. This form of nonreductive physicalism is compatible with a functionalism that claims only that some, but not all, mental properties are functional.

1.6. BEHAVIOR

A central issue in the mental causation debate is what it is that mental states are supposed to causally explain. The nature of the explanandum is important because claims of causal efficacy are always implicitly relative to a given effect. One class of arguments discussed in Chapter 2 – Arguments from Narrow Explananda – turns on just this issue. However, discussions of mental causation often employ a single term – “behavior” – to refer to differing, and competing, views of the main type of explanandum that is of interest here. In this section I will clarify different views of action and behavior and different uses of the term “behavior”.

Pretheoretically, a complete accounting of the types of causal relations in which mental states may be involved would include (a) mental phenomena that cause physical phenomena (pains that cause twitches), (b) physical phenomena that cause mental phenomena (twitches that cause embarrassment), and (c) mental phenomena that cause mental phenomena (one thought leading to another). Assuming materialism is true, the mental phenomena in each case are not completely distinct from physical phenomena. But the same goes for physical phenomena: a fourth type would include (d) those relations in which intentionally described inputs that cause mental states (seeing an

⁷¹ Kim (1998) explores these issues. See also his (1993).

elephant causing me to think about elephants) and mental states that cause intentional actions, which are not mere bodily motions. What actions are, of course, is a matter of debate, which I discuss below.

A complete account of mental causation should show how all four types of mental-physical transaction are possible. However, my aim here is limited to addressing in detail only the final category, which is the paradigm case of intentional causation, if there is any. Thus, an account which can explain mental causation in these cases will be considered satisfactory if it leaves open the possibility of accounting for the other types.

1.6.1. "Behavior"

The term "behavior" is widely used in the literature to refer to that which intentional states typically explain – e.g. Clyde's going to the fridge to get a beer. However it is also used to describe what organisms that do not have full-fledged intentional states do – e.g. mating behavior. Human intentional action is thus treated pretty universally as a particular species of animal behavior, such that "behavior" is used to refer to a frog's fly-catching as well as Clyde's cheating on his tax forms. "Behavior" is defined so as to cover the whole range: Millikan (1993), for example, defines behavior as the functional form of an animal's output, while Dretske (1988) defines it as a process of internally-caused motion, where the internal cause may but need not be a full-fledged intentional state.⁷² Thus, "behavior" may legitimately refer to anything from reflexive behaviors to complex intentional actions.

⁷² Millikan denies any principled distinction between an organism and its environment – which implies that the organism's behavior, insofar as it involves causal relations between the organism on the one hand and the environment on the other, need not be limited to intuitively internal motions. However, while Millikan claims (1993: 137) that "the functions that define behaviors often reach very far out into the environment",

This assimilation is justified from a naturalistic point of view. Unfortunately, it has helped perpetuate an important source of obscurity in discussions of the causal role of psychological states. It is not always clear when a philosopher is using “behavior” (or even “action” or “output”) to refer to bodily motions or to intentional actions. For example, Pettit (1986) begins a discussion of psychological explanations of “behavior” in which what is being explained seems at first to be the ordinary notion of action, but turns out to be bodily motion.⁷³ Others coin terms that include “behavior” or “action” as a part, but in fact use these terms to refer to bodily motion: examples include Stich’s (1983) “autonomously described behavior”, Block’s (1990) “behavioral output” and Kim’s (1982) “bodily action”.⁷⁴

This indiscriminate use of “behavior” has allowed many philosophers of mind to avoid making explicit their view of the relation that holds between bodily motion and action. The most common, if implicit, view among many philosophers of mind seems to be that bodily motion and action are token-identical, such that “bodily motion descriptions” and “action descriptions” are true of the same token events. This view mirrors standard non-reductive physicalism for mental states. The term “behavior” is then used as a way to refer to these tokens in what seems to be a neutral manner.

However, many action theorists consider this view incorrect.⁷⁵ While this view sounds comfortably Davidsonian, Davidson holds that only bodily motions with an intention in their etiology can have action descriptions. This widely accepted

she does not appear to endorse the more radical claim that the functions are constitutive properties (in the sense of 1.3.1), as opposed to just descriptions of external changes. Dretske is not clear on this point.

⁷³ Pettit (1986: 34) appears to use the term “action” ambiguously to refer to human action and the sort of action involved in the principle of “no action at a distance”.

⁷⁴ Block (1986: 668) also appears to assimilate “behavioral effects” to “impulses in motor-output neurons”, such that after using these phrases his subsequent reference to “jumping behavior” is ambiguous..

⁷⁵ Hornsby (1980: 2) calls it “*prima facie* implausible” to identify (a token of) my arm’s moving with (a token of) my moving my arm.

characterization does not entail token-token identity. If C is the intentional cause and M the resulting bodily motion, a token action can be identified with M, with C, or with the C → M process. All three of these positions are occupied, not just the first. A significant number of action theorists think bodily motion is a constituent of an action along with the intention that causes it. Dretske (1988), following Thalberg (1977), supports this view. Others, such as Hornsby (1980), identify the action with the intention (or with a mental event of trying), in which case a bodily motion is not even a *part* of an action. On this view, a bodily motion sustains an action description in virtue of the fact that it is caused by an action. Moreover, the claim that the token action is M is ambiguous: for to say that an M with a particular kind of cause is an action (or can have an action description) may mean that M is a functionally defined token – and like all functionally defined tokens it is not to be identified with the movement that satisfies the role.

The problem with the implicit token-token identity view, besides the fact that there are no arguments for it that I know of, is that it prejudices the case against intentional causal efficacy. Properties (of causes) are never causally efficacious *simpliciter*, but always in relation to some property of the effect, even if the effect-property is not made explicit.⁷⁶ If it is presupposed that what we are trying to explain in psychological explanations are *really only* bodily motions, then difficulties for showing how contents can be causes are built into the way in which the problem is framed.⁷⁷

⁷⁶ For example, Block's (1990) argument in which a metal rod is connected to a bomb and heated is supposed to show that the rising thermal conductivity of the rod is causally efficacious but its rising electrical conductivity is not. The argument fails because it assumes that the only relevant effect is the bomb's going off; if the rod is also part of a circuit which closes and turns on a light when the electrical conductivity rises, then the electrical conductivity is causally efficacious – relative to the light.

⁷⁷ Dretske (1988: 36) is well aware of this: "[I]f one uses this model [in which behavior = output] to think about human and animal behavior, great harm can be done. One can easily be misled into thinking that the cause of behavior is necessarily the cause of output. And once this confusion is in place, one will have no choice but to identify causal explanations of why we *do* the things we do with causal explanations of why

Following most writers on this topic, I will use “behavior” in the ethological sense in which includes but is not limited to human intentional action, but unlike many writers I do not assume that actions and bodily motions are token identical. When introducing other philosophers’ views, however, I will use their terminology initially even if this means using “behavior” as if it denoted bodily motion. I will make this clear in the text.

1.6.2. Intentional Action

Human actions are behaviors which have an intention as a cause. Intentions are psychological states which may be composed (reductively or not) of beliefs and desires or perhaps may be what desires become in certain circumstances. A creature that does not have such psychological states cannot perform actions by this restrictive definition. But even if it is widened to include lesser cognitive states, the class of actions will remain more restricted than that of behavior as a whole due to this requirement. Amoebas behave, but they don’t act; they don’t have the cognitive wherewithal.

“Basic actions” are, in Goldman’s terms, actions that “generate” higher- or same-level acts that we believe would be generated by the basic acts.⁷⁸ For example, my extending my left arm is a basic action that generates the higher-level action of my calling a taxi. I perform the basic action because I have certain beliefs about what higher-level (or same-level) actions will be generated by it. Basic actions are typically described in terms of the bodily movement that is performed: my extending my arm, as opposed to

our body *moves* the way it does. One will, in other words, have succeeded in confusing psychological explanations of behavior with neurobiological explanations of motor activity. Reasons – our *thinking* this and *wanting* that – will have been robbed of an explanatory job to do.”

⁷⁸ See Goldman (1970: 63-72). The main idea of a basic action is that it is a token of a basic act-type, where a basic act-type is the sort of property that is “exemplifiable at will” by an agent in standard conditions – a property for which it is true that if S wanted to do A at t, she would do A at t. Raising one’s hand is a basic act-type (for most of us), while wiggling one’s ears is a basic act-type for only some of us. Goldman’s other examples include moving one’s finger, bending one’s knees, shrugging one’s shoulder and so on.

my calling a taxi. However, basic actions are still actions, and are not to be confused with bodily motions. My extending my arm is a basic action, while my arm's extending is a bodily motion and not an action at all.

Although all actions have an intention in their etiology, there is a second sense of "intention" in which only some actions are intentional and others accidental: S does A intentionally only if S intends to do A (or has an intention to do A).⁷⁹ My reaching for the glass was intentional, my spilling the wine was not. Unintended actions still have an intention as cause, but they are actions under descriptions that cannot also be used to describe the original intention. This sense of "intentional" is a type-level distinction between different ways to characterize actions. Under a fine individuating scheme, my moving my arm is a different action from my spilling the wine; under a broader scheme, these are the same action under different action descriptions.

These differences in the individuation of actions do not matter for my purposes. But there is a distinct individuation issue that does matter. This is the question of whether at least some actions (like at least some thoughts) are externalistically individuated in the sense of robust realism, in which the constitutive properties of the tokens are their action properties, not their muscular properties. The idea that behavior is defined relative to an environment would seem to support a robust-realist construal of action individuation, although neither Millikan nor Dretske are entirely clear on the issue. The token-identity view, of course, is a form of weak realism about actions: a token action just is a bodily motion, otherwise described. This difference matters in the same way that the difference between robust and weak realism about mental states does. A defense of intentional

⁷⁹ This is the "Simple View" defended by Adams (1986) and Devitt (1996).

causal efficacy is likely to be stronger if the effects of intentional causes are not reducible to bodily motions at the token level. However, I will leave this issue open for now.

1.7. RELATED TOPICS

In this section I will briefly present topics that impinge on mine in some way. My intention is just to clarify the main problem in each area that affects my topic, show how it is related, and indicate how these problems might be addressed. The discussions will thus be fairly incomplete.

1.7.1. The Status of Folk Psychology

As a first approximation, folk psychology is, as Horgan and Woodward (1991) put it, “a network of principles which constitutes a common-sense theory of how to explain human behavior.” This is a first approximation because while some philosophers, like Churchland (1981), characterize this “network” a system of laws such that the intentional properties or kinds that do the explaining are defined by their place in the network, others, like Millikan (1986) or Schiffer (1991), deny that folk psychology is a theory of this type and that it explains by means of laws at all.

However, on anyone’s view among the most central explanations folk psychology provides are those that invoke mental tokens that are explanatory of behavior in virtue of their intentional contents. If someone desires that *p*, and believes that doing *y* will satisfy that desire, then, *ceteris paribus*, that person will do *y*. More generally, in Fodor’s (1994) terms, “an intentional explanation is one that subsumes psychological states by reference to their intentional contents”. A psychology that explains behavior by citing intentional

states that for some reason do not qualify as beliefs, desires or so forth would be intentional psychological but not folk psychological.

The relevant problem for my topic raised by the debate over folk psychology is that of intentional realism. Critics of folk psychology often hold some form of intentional irrealism (which includes eliminativism, instrumentalism and, arguably, reductionism) as their fundamental philosophical commitment, and their attacks on the legitimacy of intentional causal explanation aim to support this commitment. Much of this hostility toward the intentional is inspired by Quine (1960), who argued that the propositional attitudes do not deserve a place in an austere scientific ontology, and that intentional talk is merely a “dramatic idiom”.

For example, Churchland (op. cit.) argues that folk psychology is a false theory; we ought to presume, therefore, that its terms do not refer. On his view, “active coherence” with what we know in other sciences is “the final measure of any theory”. Reductionism would provide this coherence, but reduction is unlikely due to folk psychology’s explanatory impotence, its millenia of stagnation, and the fact that its categories don’t match up with those of the neurosciences. The only other reasonable choice is eliminativism.⁸⁰

Horgan and Woodward (op.cit.) reply that if reductionism fails, Davidson’s non-reductive physicalism is another viable naturalistic alternative. This response is weak if only because Davidson’s view has been roundly criticized for rendering mental properties epiphenomenal. They also note that folk psychology has not been stagnant; cognitive science and neuroscience actively seek the neural bases of memory, visual experience,

⁸⁰ Stich (1983) also insists on the alleged inadequacy of folk psychology. Other arguments by Stich are directly relevant to my topic and will be discussed in later chapters.

imagery, language processing, and so forth, all of which are folk psychological notions. I would add that whether the localization of cognitive functions in various areas of the brain constitutes a reduction of mental states to brain states is an open question. Showing that there are lower-level mechanisms that can carry out certain functions does not entail the reduction of higher-level functions to lower-level mechanisms.

1.7.2. Non-Reductive Physicalism

Physicalists hold that in some sense everything that exists is physical, where what “the physical” is is left to physics (or perhaps the physical sciences) to determine. Reductive physicalists hold that everything is physical or can be reduced to the physical, while non-reductive physicalists deny that mental properties can be reduced to (or identified with) physical properties. Some non-reductive physicalists also deny token-token identity, although that is a minority view. Non-reductive physicalists also differ regarding the type-level relation that replaces identity – supervenience and realization are the usual relations suggested – but these differences will not be relevant here. That said, for convenience I will usually use the term “supervenience” to characterize that relation.

The relevant, and very large, problem posed by the debate over non-reductive physicalism is whether there is a version of it that escapes epiphenomenalist worries. The version that typically has raised these worries is what I have called weak realism: token-token identity plus the denial of type-type identity. If the arguments are correct that all supervenient properties (those posited by higher-level theories and not reducible to lower-level properties) are causally inert, then intentional properties, as a special case, are causally inert.

This issue has been pursued most prominently by Kim in the form of the problem of causal exclusion. According to Kim, it seems that the causal powers of supervenient properties are preempted or excluded by the causal powers of their supervenience bases, given the causal closure of the physical (the claim that all physical events have physical causes.) Some theorists deny that there is a problem: if mental states really are token identical to physical tokens, then nothing gets excluded.

Pereboom (2002) (discussed in Ch. 4) also suggests that exclusion is avoided even when a mental state and its causal powers are wholly constituted by, but not identical to, a microphysical state and its causal powers. He argues that if (as Kim holds) there is no causal competition when mental and microphysical tokens are identical, then there is also no causal competition if mental and microphysical tokens merely constitutionally coincide (i.e. are actually constitutionally coincident). For the difference between identity and actual constitutional coincidence without identity is that in the latter case there is some possible world where there is no constitutional coincidence. But this possibility doesn't seem to introduce explanatory competition.

Another way to counter Kim's argument might go as follows. Kim argues, plausibly, that supervenience is "not happily construed" as a causal relation, since the instantiations of the related properties are instantaneous while the cause-effect relation is not.⁸¹ But this means that there are two kinds of sufficing involved in the case: *superveniently sufficing*, in which the presence of a property of one kind entails that a property of another kind is present, and *causally sufficing*, in which the presence of one property is a sufficient condition, in the circumstances, for the subsequent appearance of

⁸¹ See Kim (1998: 44). I should note that I have a longer paper on this topic, of which the argument to follow is just a part.

the second. In a given case of causation between a mental property M_1 and another mental property M_2 , there is no reason to think that the physical base P_1 will causally exclude M_1 from producing M_2 by causing P_2 . *Pace* Kim, M_1 cannot cause M_2 *by* causing P_2 , for while P_2 superveniently suffices for M_2 , it does not causally suffice for M_2 . And if M_1 need not go via P_2 to cause M_2 , then there is no reason to think that M_1 will be preempted by P_1 as the cause of P_2 .

1.7.3. Semantic Naturalism

The relevant problem of semantic naturalism is whether it is possible to give an account of semantic properties in which they can be shown to be part of the natural, material world. The project of naturalizing semantics is extended to that of naturalizing content by the usual assumption that intentional and semantic phenomena – thought and language – are relevantly similar. The “naturalism” consists in the claims that, in principle, semantic properties can be explained in terms of non-semantic properties (or that semantic facts obtain in virtue of non-semantic facts), and that semantic phenomena are suitable for empirical investigation.⁸² The first claim is intended to be neutral between the ideas that semantic properties supervene on non-semantic properties and that they are reducible to non-semantic properties. If naturalism is not possible, there can be no account of the causal efficacy of content that is not dualistic in either a Cartesian or a Leibnizian (“pre-established harmony”) sense. Either we would have to admit that there is an additional causal force in the universe that physics cannot ultimately account for, or we would have reason to be eliminativists about intentionality.

⁸² See Loewer (2002) and Devitt and Sterelny (1987) for explanations of semantic naturalism.

Theories of content try to state the conditions in virtue of which a given mental state has the specific type of content that it does. Those theories that rely on only naturalistic relations – such as physical causal relations or biological-evolutionary considerations -- in giving this account are naturalistic.⁸³ These would include the theories put forward by Dretske, Fodor and Millikan, among others.

On Dretske's "information-theoretic" view, the content of a belief is the information it carries under certain circumstances. For a belief to carry a certain type of information is for there to be a natural regularity of some sort such that when a belief of that type occurs, that circumstance obtains. On Millikan's "teleological" view, the content of a belief is the information that it is the belief's biological function to carry – the information it was naturally selected to carry. There need be no regularity between the belief's occurring and the occurrence now of a certain circumstance, since biological functions are properties that increased the fitness of ancestors, whether or not they continue to do so or whether they do so often or with regularity. On Fodor's "asymmetric dependence" view, the content of a belief is the information it carries under certain content-fixing circumstances. Content-fixing circumstances are content-constituting causal relations between a certain circumstance and an instance of a given belief type. The asymmetry lies in the causal relations: a belief of type A can be caused by causal relations that are not content-constituting as well as those that are.⁸⁴

⁸³ Naturalism is not synonymous with externalism. A naturalistic account would also have to account for the logical connectives and other internally individuated contents. However, most naturalists focus on giving an account of externalistically individuated contents.

⁸⁴ See Loewer (unpublished) for a summary of naturalization projects and the problems that each face.

One issue often raised against these proposals is that of the so-called normativity of content.⁸⁵ There seem to be two issues here. One is the claim that it is constitutive of a concept that it be applied only to certain objects and not to others, or that it is correctly used in a certain way and not others. The concept “cat” *ought* only to be applied to cats. This seems to mean that intentional contents somehow involve an “ought” that cannot be reduced to a naturalistic “is”. A related, but distinct, claim is that our ascriptions of content are constrained by norms of rationality and charity. We ascribe contents in accordance with idealizations about how people rationally ought to think.⁸⁶

As far as I can tell, the second claim has no force against *naturalism*. That ascriptions of belief are normative does not entail that contents are normative, unless one confuses the uses of content-ascribing sentences with the mental contents themselves. Criteria of rationality and charity affect our willingness to ascribe beliefs and our patterns of doing so, but this gives us no reason to think that even irrational or incompletely ascribable beliefs can’t be *explained naturalistically*.

The first difficulty is harder to assess, and I won’t pretend to be able to do so here. It may be a matter of how robust a conception of “ought” is being assumed. Rey (1997) distinguishes between descriptive and prescriptive normativity (or descriptive vs. normative idealizations) in response to the rationality issue, but his distinction seems to apply to this problem as well. The first sort involves explanatory idealizations that are found throughout the sciences, while the second involves genuinely normative ideals which are standards of judgment that may never be instantiated. The former idealizations

⁸⁵ Another problem often raised is whether such accounts can overcome indeterminacy. This is not a problem for naturalism as such, since indeterminate contents can be naturalistic in that no non-natural notions are invoked to explain them.

⁸⁶ This is Davidson’s fundamental argument for anomalous monism, as he freely concedes in his “Replies” to Suppes (1985). See Rey (op. cit.: 74-77).

will say (e.g.) how gases ought to behave, but this “ought” is not a prescriptive ought. If the sort of “ought” in content is no more than this, then there is no threat. It is not obvious that the stronger “ought” is involved.⁸⁷

1.7.4. Models of Cognition

Currently, there are three main competing models of cognition: orthodox or classical computationalism, connectionism and dynamic systems models. According to “classical” computationalism, minds are a special kind of computer in that cognitive processes (or, at least paradigmatically, those patterns of thought found in deductive inference) are rule-governed mechanical manipulations of syntactically-structured symbols in virtue of their syntax or other internal (sometimes called “formal”) properties. Connectionism explains cognition in terms of variably-weighted causal relations in a complex network of representational nodes that, unlike the symbols in the orthodox model, do not have syntactic structure and so do not interact in virtue of syntax.⁸⁸ Dynamic systems models explain cognition in terms of temporally continuous processes that are described by differential equations governing the relations among the variables and parameters that define a given system. Unlike the classical model, the system evolves

⁸⁷ See also Horwich’s (1998) distinction between the normative implications of a meaning property (e.g. regarding how it ought to be used; the fact that ‘dog’ ought not to be applied to cats) and the intrinsic normativity of a meaning property. The fact that meaning properties have normative implications does not entail that meaning properties themselves are normative.

⁸⁸ Fodor and Pylyshyn (1988) claim that connectionism and classical computationalism are compatible in that a classical computational system can be implemented in a connectionist architecture; they are theories about different cognitive levels. Connectionists usually see their view as a competing account of cognition at the same level of explanation.

without discrete representations or discrete changes from one overall state of the system to the next.⁸⁹

Relative to my dissertation topic, the relevant issue is to show how contents can play a causal role within the explanatory and conceptual framework posited by each model. The models may not be equal in this respect. For example, according to van Gelder (1995), the dynamic systems model holds that a cognitive system is not encapsulated in the brain, but includes (in the form of numerically measurable parameters or variables) changes in the body and environment. This holds out the possibility that changes in referential relations, *to the extent that they are measurable or quantifiable*, can figure in the dynamics of a cognitive system. On the other hand, Block (1990) has argued that it seems to follow from the classical model that the intentional contents of mental representations, which are not manipulated by the system, cannot be causes.

I will focus (in Ch. 4) on the classical model, since the most worked-out version of my dissertation topic as it relates to cognitive science has been expressed in its terms. In particular, Fodor (1994) casts the problem in terms of explaining how computational mechanisms can implement intentional psychological laws if the internal properties of mental representations do not suffice to encode the contents that figure in those laws. If the mind is a kind of computer, then a good explanation of why the psychological laws hold is that they are implemented in lower-level laws that subsume these events in virtue of their internal properties. But if no internal (physical or syntactic) property can suffice for wide content properties, we have no good explanation of the psychological laws, and hence no good explanation of intentional causation.

⁸⁹ See van Gelder (1995) for a discussion of all three views and Rey (op. cit) for classical and connectionist views.

1.8. CHAPTER SUMMARY

In this chapter I have tried to introduce and clarify the important concepts and terms that are used in discussions of my topic: intentional content, externalism, belief, mental event, cause, causal explanation, relational properties, behavior. With these clarifications assumed, my topic may be simply, if technically, stated as follows. Beliefs and other mental states with externally individuated intentional contents figure as causes in virtue of their contents in many successful psychological causal explanations of behavior. The question is whether these contents play genuine causal roles. Starting in the next chapter, I will discuss particular arguments that answer this question in the negative, and the main responses to them.

CHAPTER TWO: SOME INITIAL ARGUMENTS

2.1. INTRODUCTION

2.1.1. Review

In Chapter 1 I introduced the problem of externalism: roughly, how can wide contents be causes of behavior? I motivated my thesis by noting that wide contents are among the properties we ordinarily cite to explain our behavior. I will be supporting the claim that beliefs and other mental states with widely individuated intentional contents play genuine causal roles in virtue of their contents in psychological causal explanations of behavior. In abbreviated terms, the question is whether wide contents are causal or are causal powers or play causal roles. I will answer “yes”.

I also set the stage for looking at the problem in more detail by defining the concepts and background assumptions that explicitly or implicitly play a role in the literature on this topic. Doing this involved discussions of externalistic individuation, intentional realism, causation and causal explanation, relational properties and behavior. In this chapter I begin examining specific arguments against the causal efficacy of externalistic mental states. Arguments that such contents *do* not play causal roles are straightforward; a reason is given and the conclusion is supposed to follow. Arguments that they *ought* not to play causal roles follow indirectly from arguments to the conclusion that a certain adequacy condition on a genuinely scientific cognitive psychology is justified, plus the claim that intentional psychology and the wide states it posits do not meet this adequacy condition. Although it is always possible to respond to prescriptive arguments by rejecting the recommended adequacy condition, these arguments tend to force the defender of intentional causes to justify her claim that

ordinary intentional causal explanations are genuinely causal. Supporters of the prescriptive approach typically grant that we may continue to provide psychological explanations in which contents figure as causes. But they often take their arguments as having shown that these claims should not to be taken any more seriously than the claims that the sun rises and sets (see 1.4.3).

The arguments of this chapter are preliminary to what I consider to be the more fundamental Arguments from Non-Locality, addressed in Ch. 3. Here I present first a brief argument from Fodor which raises the issue of the role that the individuation conditions of a property play in the causal explanations that cite that property. I then turn to Arguments from Narrow Explananda, which rely on certain claims about what the explananda of psychology *are* or *ought to be* for the purposes of causal explanation. The final grouping includes Fodor's Cross-Context and Modal Arguments, which attempt to distinguish content individuation from causal-power individuation.

In the Arguments from Non-Locality, the opposition to contents as causes is tied directly to their extrinsic nature. That is the main reason why I consider these arguments fundamental. Also, the arguments in this chapter seem to rely implicitly on the causal intuitions that drive the Non-Locality arguments. In Ch. 4, I present a solution to the problem, showing how extrinsic causal powers are perfectly acceptable in science and how wide contents can be genuine causal powers within a scientific psychology.

2.1.2. Extrinsic Causal Powers

Before presenting arguments against extrinsic causal powers, it might be useful to have in mind some examples of extrinsic properties that *prima facie* are robust causal

powers. This will help counteract the impression given by widely known examples in the literature (many of which will be rehearsed in this and the next chapter) that the very idea of such a thing is absurd.

Illocutionary acts. Fairly close to the usual cases under discussion in this context are the properties of sentences uttered in illocutionary acts.⁹⁰ Whoever occupies the role of being the President of the United States has the power (in a national emergency) to declare war, a power which no other individual, however physically identical, has. An utterance of “I formally declare that the United States is at war” is an illocutionary act of a special sort – an event that causes troop deployments, arms production and so on – only when uttered by an individual in that position. Neither I nor a molecular duplicate of the president who coalesces in space can perform that illocutionary act, no matter how identical the movements of our mouths, the form of words we utter or the semantic meaning of these utterances. Similarly, a judge can condemn a felon by saying “I sentence you to 10 years in jail without possibility of parole”. Molecular duplicates who are not judges cannot perform that act and bring about those effects, even if they perform the same bodily motions.

In both these cases, the individuals have certain causal powers in virtue of the social positions they occupy, powers which molecularly identical individuals who are not in those social positions lack. As a result of those social relations, some of their utterances are events (declarations of war, sentencings) that do not otherwise occur; no utterance of mine will ever be a declaration of war or a sentencing (unless I become U.S. president or a judge). It is implausible to deny that these differences are causal

⁹⁰ I thank Prof. John Greenwood for suggesting this example.

differences and that the relations in which certain individuals stand are essential to these causal differences.

Chiral molecules. Another case is the phenomenon of chirality or enantiomerism in chemistry.⁹¹ Molecules with the same composition but different structure are called isomers. For example, ethanol ($\text{C}_2\text{H}_5\text{OH}$) and methoxymethane (CH_3OCH_3) have the same atoms, but they are arranged in different functional groups. One result is that ethanol is highly reactive and methoxymethane relatively unreactive with sodium.⁹² Among the isomers are the stereoisomers, which have the same functional groups, but they are arranged differently in space.

Among the stereoisomers are the optical isomers, which are asymmetric and handed (left or right). The “optical isomer” name comes from their characteristic of rotating the plane of polarized light either to the left or the right. This feature played an important historical role in the discovery of chirality: in 1848, Louis Pasteur was led to isolate enantiomers after finding differences in optical activity in tartaric acids deposited by maturing wines.⁹³ Today, these molecules are usually called chiral molecules (from the Greek *kheir*, “hand”) or enantiomers. Such molecules lack a plane of symmetry; they are mirror images of each other which cannot be superimposed, or rotated to coincide with each other, in a given space. Their internal identity is easiest to imagine by analogy

⁹¹ See Brown (1990), Crossley (1995), Jacques (1993), Janoschek (1991), and Le Poidevin (1994),(2000). Chirality is also present in quantum mechanics, in the rotatory motion of subatomic particles, e.g. whether they spin to the left or the right. I have not pursued the question of the effect of the direction of spin on the causal powers of basic particles. But there may well be further evidence of extrinsic causal powers to be found at that basic level.

⁹² Structural isomers differ in how the atoms are arranged, including in particular the arrangement or types of functional groups, or closely connected groups of atoms that are responsible for characteristic reactive properties of molecules. Ethanol and methoxymethane are structural isomers.

⁹³ Pasteur noticed that the major component in deposits of tartrate salts, produced by tartaric acid, rotated the plane of polarized light to the right (clockwise), but an apparently identical minor component had no such effect. He found that the original acid was a racemate or mixture of R- and L-tartaric acid (which also comes in a meso, or symmetric, form) and traced the difference in optical properties to the crystallization of two mirror-image forms of sodium ammonium salt. (See Crossley 1995: 4).

with a “b” and a “d”: these letters cannot be superimposed in a 2D space, but they could be so translated in a 3D space.

Chirality is of enormous importance in general (that is, outside of the context of this dissertation), because life as we know it is chiral and because different enantiomers have different causal powers:

The concept of chirality in chemistry might be merely an academic one if it were not for the stunning fact that most of biochemistry is chiral. Proteins, DNA, amino acids, sugars and many natural products such as steroids, hormones and pheromones possess chirality. ... [M]ost biological receptors and membranes are chiral, many drugs, herbicides, pesticides and other biological agents must possess chirality for binding and action to occur. Most medicines such as ibuprofen, and other bioactive compounds, such as nicotine are therefore chiral. ... Positron emission tomography scans of Ritalin in the brain reveal that the D [right-handed] form concentrates in the striatum, whereas its mirror-image form distributes non-specifically over the brain. The list of molecules having differing biological activities for right- and left-handed forms is extensive. (Hicks 2002: 2-5)

Chiral molecules share many important physical properties (except for rotating the plane of polarized light) as well as chemical properties, but typically differ in their reactions with other asymmetric molecules – that is, in “chiral environments”. While some enantiomers do not appear to differ in their causal powers in chiral environments, very many do – leading some chemists to advocate labeling all racemic (mixed-enantiomer) drugs as 50% impure. Chemistry has begun to classify enantiomers as distinct chemical species, just as ethanol and methoxymethane are considered distinct.

The causal powers of enantiomers differ in various ways. Some differ significantly in their potency in producing the same effect, some have qualitatively different effects, and sometimes both forms in a racemic mixture contribute to a single desired effect. (There may be no hard and fast difference among these: in many drug studies an “inactive” or “weakly active” enantiomer can be one that does not have the

desired effect to a significant degree.) The first class includes β -blocking drugs, such as propranolol, whose left-handed form does all or most of the heart-relaxing activity for fighting heart disease.⁹⁴ An example of the second class is the antihypertensive agent α -methyldopa, which is active only in its left-handed form.⁹⁵ The third class includes indacrinone, a diuretic marketed in a 4:1 ratio of right- and left-handed forms to obtain an optimal effect.⁹⁶ Less exotically, yeast ferments *d*-fructose (dextro or right-handed) but not *l*-fructose (levo or left-handed), and *l*-nicotine is much more toxic to human physiology than *d*-nicotine.

The most famous case of chirality is thalidomide, which was originally marketed in racemic form. After its withdrawal from the market for causing fetal abnormalities, follow-up studies tied its toxicity to the left-handed enantiomer, others traced it to the left-handed form of one of its metabolic products, and yet later studies found that the different forms could convert into one another in the blood.⁹⁷ Although the case for causal differences between thalidomide enantiomers has been inconclusive, the notoriety of the drug drove home the point that chirality could no longer be ignored in drug design. As a result, pharmaceutical companies now need a specific reason to develop racemates rather than single-enantiomer drugs.⁹⁸

⁹⁴ See Brown (1990: 5-9). As another example, both forms of verapamil (Calan, Verelan, Isoptin), a calcium channel blocker used to treat angina, have similar coronary vasodilatory effects, but the right-handed form is 5 to 10 times less potent than the left-handed form.

⁹⁵ Brown (1990). The enantiomers in the weight-loss drug fenfluramine, which was sold as a racemate but pulled from the U.S. market in 1997 due to adverse affects on aortic valves, also appear to differ in their effects: the left-handed form causes antidopaminergic (dopamine-blocking) and sympathomimetic (increased heart beat) effects that the right-handed form lacks (Crossley (1995: 17).

⁹⁶ Crossley, *op.cit.*, 15.

⁹⁷ Crossley, *op. cit.*, 180-181.

⁹⁸ This has opened up new marketing opportunities for single-enantiomer drugs to compete with established racemic drugs. An example (from Crossley, *op. cit.*, 16) is in the field of serotonin reuptake inhibitors, where SmithKline Beecham and Pfizer have marketed single-enantiomer forms against a racemic form sold by Lilly.

Chirality is important in this context because chiral differences are ineliminably extrinsic. One argument for their extrinsicness stems from the fact that if two enantiomers cannot be rotated to coincide, they must be in a space of certain dimensions. As noted above, if a “b” and a “d” cannot be superimposed, then they must be in a 2D space and they are distinct relative to each other (right- or left-handed). Chiral molecules in our ordinary 3D space are analogous, only we cannot picture a 4D space in which they might be rigidly moved so as to perfectly coincide. So whether an object is enantiomorphic depends on the global geometrical properties of space, and whether it is left- or right-handed depends on its relation to other asymmetric objects in that space.⁹⁹ In terms of the definition of extrinsicness I defended in 1.5.1, a molecule’s being an enantiomer entails that it exists in a 3D space, and its being right- or left-handed entails the existence of a left- or right-handed counterpart in that space.

The objection might be raised that the differences between chiral molecules and their causal powers are due to an internally specifiable feature, some as-yet-unknown intrinsic structural property. For example, it has been suggested that the chirality of molecules may be traceable to the violation of parity (space inversion or mirror symmetry) in the electroweak forces governing β -decay of unstable atomic nuclei.¹⁰⁰

⁹⁹ This treatment of enantiomorphism is from Nerlich (1994), and is discussed, along with asymmetry and handedness, in Le Poidevin (1994) and (2002). Kant famously argued for the reality of absolute space from the claim that only by positing absolute space could we explain why, in a universe containing just one hand, that hand is either a left hand or a right hand. Would a lone hand in a universe that contained nothing else be right- or left-handed? Kant felt it had to be one or the other, but Le Poidevin and others seem to agree that it would be neither. As will be noted below, chemistry sides with Le Poidevin et al.

¹⁰⁰ See Latal (1991) and Liboff (2003: 177). Parity violations are violations of the assumption that the laws of nature do not distinguish between left and right – i.e. that mirror symmetry is conserved in all interactions. Violation of the law of conservation of parity was first observed in 1956 in the β -decay of a neutron, when it was found that the electron always came out spinning in a left-handed sense and the anti-neutrino in the right-handed sense. In principle, the handedness of a fundamental particle with mass can be reversed by bringing it to rest and accelerating it in the opposite direction without changing its spin. However, the handedness of a massless particle (such as an electron neutrino) cannot be changed; it always travels at the speed of light and cannot be stopped.

This failure of a basic conservation law appears to be due to the fact that neutrinos have only left-handed spin and, as far as we know, experience only the weak interaction. So the objection might be that chirality has an intrinsic explanation at the level of elementary particles, even though enantiomers are identical at the level of molecular structure.

But this would not make chirality intrinsic. For whether a molecule is chiral at all (and not just why it is left- or right-handed in particular, *given* that it is chiral) appears to depend in part on the kind of environment that it is in. Some studies have shown that interaction with a chiral environment can induce chirality in molecules that would not normally be considered chiral, such as propane.¹⁰¹ For example, some compounds (N,N-diisopropylarylglyoxamides) can be induced to crystallize into either right- or left-handed (or racemic) forms depending on the appropriate conditions of crystallization:

The importance [of this experimental result] for molecular recognition is that these achiral molecules self-organize into a chiral environment, and that interactions in this chiral environment are enough to differentiate between otherwise equivalent groups. These considerations are an effective demonstration that achiral molecules in a chiral environment are chiral and, by extension, symmetric molecules in an asymmetric environment are asymmetric. (Crossley, 10-11)

This finding implies that the reductionist intuition that chirality is an environmentally invariant feature may just be an artifact of the fact that we live in a highly chiral world, one in which chiral molecules as we know them exist in and maintain their chirality. We may well imagine that the chiral nature of a molecule is context-invariant even though its handedness and non-superimposability entail certain facts about its actual environment. (And if that is right, then so much the worse for a definition of “extrinsic” that classifies chiral molecules as extrinsic.) But even if there is an elementary-particle explanation

¹⁰¹ SEe Crossley (1995: 2-10). He also notes that it also depends on such features as molecular size and the possibility of conformationally mobile molecules, in which temperature-dependent conversion of one chiral form into another can occur (via rotation about bonds).

available in the case of each and every chiral molecule, chirality cannot be intrinsic if a molecule's being chiral depends essentially on its being in a chiral environment. This dependence rules out the possibility that chirality can be completely explained in terms of *intrinsic* properties (although it does *not* rule out an explanation in terms of elementary particles and their interactions).¹⁰² This issue will be addressed again in Ch. 4, where I discuss arguments to the conclusion that molecular structure in general is extrinsic.

Besides the causal-power differences between chiral molecules, the phenomenon of chirality is important in this context because it illustrates an important way in which Twin-Earth (or replica) cases are scientifically underdescribed. Twin-Oscar could be exactly like Oscar in every internal respect – they can be, as it is often said, identical in the atoms they contain and their arrangement. But if Twin-Oscar (and Twin-Earth) are mirror images, *nothing* can be concluded *a priori* about whether Oscar and Twin-Oscar have the same causal powers. (See, for example, Stich's Replacement argument below.) To the contrary, from an empirical standpoint that they probably would *not* have the same causal powers. The same goes for claims about the causal powers of molecular duplicates who coalesce in space or in a swamp. If we bar mirror images from counting as duplicates, then they may have the same causal powers, but these causal powers may be extrinsic. If we do not rule out mirror image worlds, there is no reason *a priori* to think that the causal powers of molecular duplicates are identical. So, to scribble in the margin of Stich's playbook: sameness of current, internal, physical states is *not* sufficient for sameness of causal powers. Or so our best chemistry says.

¹⁰² In Chapter 4 I will present quantum chemical arguments to the conclusion that molecular structure *in general* – and not just chiral structure – is extrinsic. This would make the reductionist response untenable.

In sum, there seem to be genuine causal differences ascribable to extrinsic differences. The cases of chirality show that two internally identical molecules can have quite different robust causal effects depending on their relation to each other and their orientation in space. Illocutionary acts are cases where individuals who occupy certain roles have robust causal powers via their utterances that individuals who do not occupy those roles do not have. Now we will look at arguments against wide causal powers.

2.2. FODOR'S ARGUMENT FROM THE FORMALITY CONDITION

Argument. Fodor (1980) argues for the Formality Condition as a criterion of adequacy for a scientific psychology. The Formality Condition holds that cognitive psychology ought to restrict the process- or state-types it uses in its generalizations to those that are formal or syntactic – that is, intrinsic or internal, or in any case not semantic.¹⁰³ A formally characterizable process or state token is one whose individuation conditions – the process- or state-type that it instantiates -- do not entail the existence of entities entirely distinct from that token.¹⁰⁴ Internally individuated properties clearly play a causal role in a computational psychology, for the properties of mental symbols in virtue of which they can be mechanically manipulated are supposed to be intrinsic.¹⁰⁵ The

¹⁰³ Devitt (1996: 258-265) distinguishes “formal” (brute-physical) properties from “syntactic” (internal-functional) properties, and notes that the distinction between the latter and certain proposed versions of “narrow content”, which are *semantic* properties, is vague. Fodor is thus not entirely clear as to whether the Formality Condition excludes *semantic* properties (wide or narrow) or *relational* properties (which relevantly includes only wide semantic properties). Since wide contents are excluded either way, it will do no harm to read Fodor's reference to “semantic” as equivalent to “wide semantic”. As noted in 1.5.3., functional properties are wide or narrow depending on how inputs and outputs are defined.

¹⁰⁴ In Putnam's (1975) terms, a psychological state in the narrow sense is one the ascription of which does not presuppose the existence of individuals other than the subject of the ascription; wide ascriptions do entail the existence of other individuals. This accords with the definitions of relational properties in 1.5.1.

¹⁰⁵ See Block (1990) for a clear statement of this line of thought. Whether these properties are the *only* manipulable properties is another issue, which will be raised in chapter 3.

Formality Condition forbids any non-individualistic types from playing a causal role. So if it is justified, the claim that contents play a genuine causal role would not be.

Fodor argues that a restriction to formal or internal properties is advisable because it allows us to explain, within the context of the computational model, how mental states with transparently identical wide contents (e.g. Morning Star/Evening Star beliefs) can have different behavioral effects.¹⁰⁶ He supplements this reason *for* a psychology that satisfies the Formality Condition with an argument *against* an intentional psychology that uses contents in its generalizations. I will focus on this latter argument.

According to Fodor, we ought not to pursue a “naturalistic” psychology (one that uses semantic properties in its generalizations) because we can’t feasibly get one. A psychology of this sort must specify the organism/environment relations that hold when the former is thinking about the latter. So it must define its nomological generalizations over mental states on the one hand and external entities on the other. But for these generalizations to be lawlike, we need canonical descriptions of the external relata – physical descriptions which denote constitutive properties of the external entities. This means that psychological generalizations can’t be formulated until after the basic sciences have found all their constitutive properties – for example, after chemistry has discovered that water is H₂O. It’s OK to have a research programme that puts off some questions for later, Fodor asks, “but who wants to wait *forever*?” In short, the difference between not pursuing an intentional psychology and pursuing one is “real but academic”;

¹⁰⁶ Fodor (1980: 492). Devitt (1996: 140-50) responds to this argument for an individualistic psychology by arguing that an intentional representational theory of mind (one that uses wide-semantic properties in its explanations) need not be a “direct reference” theory: on his view, contents are always referential, but there may also be additional factors (“modes”) involved in their determination. If this is right, Fodor is wrong to equate “wideness” with “transparency” and “narrowness” with “opacity”, for if all content properties are referential (“wide”), then “opaque” contents in particular are referential (“wide”).

in effect, psychology will either follow the Formality Condition explicitly or be pursued in accordance with it anyway.

Some initial remarks. The lawlike generalizations that Fodor is referring to are claims such as (1) “‘dog’ tokens are reliably caused by dogs”, as opposed to (2) “beliefs that it’s going to rain are reliably correlated with umbrella-carrying behavior”. That is, he construes psychology as a “naturalistic psychology of reference”, not as a science of behavior and cognition. Fodor never justifies this construal of “psychology”. Yet if we accept his construal for the sake of argument, his argument amounts to the claim that one must possess a complete, physically specified theory of content before one can begin doing psychology or, assuming that the theory of content is a part of psychology, that we must do that part first before we can do anything else. Either claim is descriptively false. Our lack of a completed theory of content hasn’t stopped anyone from pursuing psychology as normally construed (as a science of behavior and cognition) or from pursuing parts of psychology other than the theory of content. For example, the debate over whether the theory of vision essentially uses wide notions in its generalizations has been pursued independently of whether a theory of content for these notions is complete.¹⁰⁷ This means, first, that there is a very real and non-academic difference between pursuing an intentional psychology and choosing not to pursue one, and, second, that the threat of having to wait forever can’t be all that threatening if no one is *in fact* waiting.

¹⁰⁷ See Burge (1986) and Egan (1994). Neither doubts that intentional notions are being used (in the sense of “intentional” defined in 1.2.1) regardless of the details of how they are individuated.

Stich's Response. Stich offers two replies that rely on general considerations about scientific methodology that Fodor seems to contravene.¹⁰⁸ His first observation is that many sciences – such as descriptive botany, ethology and paleobiology -- do not seek nomological generalizations. So Fodor's demand that a scientific psychology must seek such generalizations is unjustified. An intentional psychologist's generalizations will "in some way or another *involve* those organism/environment relations ... but it is not at all clear that the psychologist [of this sort] must seek *nomological generalizations about* those reference and content determining organism/environment relations." And if there is no need to seek nomological generalizations to do science, the problem of waiting forever to get them does not arise and we have no reason to pursue a psychology that obeys the Formality Condition.

Stich's second response is that there is no reason to think that the external entities on the environment side of the content-fixing relations must be characterized in terms of basic physical properties in order for the resulting intentional notions to be adequately defined. The external entities may be respectably classified in special-science taxonomies which do not reduce smoothly to those of the physical sciences. Fodor himself has argued for the "scientific respectability" of special-science classificatory schemes which cut across those of the basic physical sciences.¹⁰⁹ So if we don't need to reduce the external relata to entities of the basic sciences, then again there is no reason to think we must wait forever to do intentional psychology. So again we have no reason to pursue a psychology that obeys the Formality Condition.

¹⁰⁸ See Stich (1983: 162-4). Stich himself argues against the use of semantic notions in psychology; see section 2.3.

¹⁰⁹ See Fodor (1975: 1-26).

Burge's response. Burge argues, also from a general scientific methodological standpoint, that Fodor conflates properties that are used in a science's explanations – its “explanatory kinds” – with properties that help define these explanatory kinds; the latter may not be part of *any* science.¹¹⁰ As we saw with Stich's second response, he argues that the special sciences can make use of unreduced intentional notions as explanatory kinds. Burge adds that sciences typically *presuppose* the individuation conditions of these explanatory kinds as background conditions. The individuation conditions need not themselves be explanatory kinds, either in psychology or any other science, nor need they be explained in a science: the individuation conditions may not figure as explanans or explanandum in the relevant science or in any science. To cite his example, physiology presupposes but does not explain the biological conditions that define what it is to be a heart, which is one of its explanatory kinds. Similarly, *bears such-and-such relation to H₂O* is not an appropriate kind for use in scientific (psychological) explanation; *has the concept 'water'* is of use in psychological explanation, but the relations that individuate it need not play any role in psychology, or any science. If these relations are presupposed and are not themselves explanatory kinds either as explanantia or explananda, there is no need to wait forever to formulate these relations in the form of lawlike generalizations in order to do psychology. So we have no reason to pursue a psychology that obeys the Formality Condition.

According to Burge, this conflation also plays a part in an argument by Fodor and Block that lampoons the externalist's position regarding explanations of behavior.¹¹¹

¹¹⁰ Burge (1989: 313-316). This conflation also plays a role in Fodor's Modal Argument (see 2.4.2 below).

¹¹¹ Burge (1989: 314); the example is in Fodor (1987:40); both attribute the original point to Block. I *think* the reference is to Block (1986: 625), although his version does not “lampoon” the externalist's position; that's Fodor's contribution. This argument is discussed further in Chapter 3.

They argue that if wide contents are causal powers, a science studying the etiology of differences in food preferences between Twins – who have the same genetic endowment and early training – would have to cite “linguistic affiliation” as the source of Twin1’s preference for brisket and Twin2’s preference for twin-brisket.¹¹² One train of thought in this argument is the idea that the relations that individuate brisket and twin-brisket beliefs are themselves part of the explanations in which those beliefs figure.¹¹³ But, Burge notes, this conflates the science’s explanatory kinds (brisket, twin-brisket) with the relations that are needed to individuate those kinds. These relations need not be cited as an explanatory kind in that science; and they can be presupposed as background conditions of the properties that *are* its explanatory kinds.

Summary. Fodor justifies the Formality Condition as a criterion of adequacy for psychology by arguing that the external relations that individuate content cannot be scientifically (lawfully) formulated until the external relata are reduced to entities in a basic science. Stich’s main response is that this demand for reduction is unjustified if the special sciences have legitimate classificatory schemes. Burge’s main response is that the individuation conditions of a science’s explanatory properties are typically presupposed by that science and need not be explained by it.

I think these responses adequately address Fodor’s argument. Burge’s is stronger because it addresses the issue of the role of individuation conditions in a science’s causal explanations. Even if (*pace* Stich) Fodor’s demands for nomological generalizations and

¹¹²Twin1 believes falsely that brisket, a cut of meat from any four-legged animal’s breast, is only from cattle, while Twin2 believes truly that twin-brisket is only from cattle. Fodor (1987: 40) concludes: “But surely it’s *mad* to say that linguistic affiliation is per se a determinant of food preference; how *could* it be?”

¹¹³The other two lines of thought in this argument fall under the broad headings of what I am labeling “Arguments from Narrow Explananda” and “Arguments from Non-Locality”. The former are discussed in this chapter, the latter in chapter 3.

reductive characterizations *were* justified, Burge's argument shows that it *still* would not follow that we would have to do these *prior* to doing psychology. For these relations themselves need not play any role in the explanations in which the wide notions they individuate figure. This in fact seems to be the case regarding the theory of vision.

I would add that Fodor's demand for explanatory priority for lower-level properties that in some sense explain properties of non-basic sciences is just as dubious as his demand for explanatory priority for individuation conditions.¹¹⁴ It is reasonable to seek lower-level implementing properties and mechanisms in psychology, but unreasonable to think that we must wait to do psychology until after the implementing mechanisms have been discovered. To the contrary, it is hard to see how we could look for a lower-level implementing mechanism of a psychological law or a lower-level property that realizes content unless we first have a grasp of that law or that content's explanatory role.

What this extension of Burge's response shows is that the fact that semantic notions involve external relations is *inessential* to Fodor's argument. The considerations he brings to bear against wide contents are quite general, and are no more acceptable in the general case. For example, if Fodor were correct, syntactic properties could not be used in psychological explanation, since we don't have nomological generalizations in physical terms for defining syntax. The assumption that we *can* get them is not the same as having them first, which is what Fodor's argument demands. But if it is OK to assume that we *can* get them for syntax, then externalistic content is *no worse off* than syntax as

¹¹⁴ Fodor (1994: 8-11) characterizes an implementing mechanism as one in virtue of whose operation the satisfaction of a special science law's antecedent reliably brings about the satisfaction of its consequent: water freezes if cooled, and the mechanism that implements this law involves changes of the molecular structure of water that cooling induces. The property relations that implementation requires are, typically, reduction or multiple realization (or some other non-reductive relation).

far as Fodor's argument is concerned. For his argument presupposes that the external relata in content-determining relations *can* be given physical descriptions; Fodor hasn't argued that such descriptions are impossible, only that we must wait forever to get them. In fact, if Fodor is right that we *must* have nomological generalizations in physical terms before we can do psychology, the fact that semantic properties involve *external* relations is irrelevant, since a formal psychology based on syntax faces the same problem.

We may conclude that the Formality Condition, which rules out using semantic properties in psychological explanations, has not been justified by this argument as a restriction on all of psychology. So Fodor has not given us a reason to change the status quo in which we cite contents as causes in psychological explanations.

2.3. ARGUMENTS FROM NARROW EXPLANANDA

Introduction. In this section I discuss a group of arguments against intentional causes that depend on claims about the proper explananda of psychology. The arguments focus on claims about behavior, which is one of the main targets of psychological explanation and the area in which intentional states most clearly play a causal role if they play any such role at all. The interpretation of "behavior" is a large part of the dispute.

2.3.1. Stich's Replacement Argument

Argument. Stich's Autonomy Principle plays a similar role in a prescriptive argument against the use of contents in psychology as Fodor's Formality Condition.¹¹⁵

¹¹⁵ Stich (1983: 164-170). The Autonomy Principle (or "principle of psychological autonomy") is first introduced, but not defended, in his (1978), where his aim is to show that ordinary psychological explanation involving intentional states violates the principle. Without a defense, many would reject the principle rather than ordinary psychological explanation.

The Autonomy Principle holds that a scientific psychology ought to restrict its generalizations to those that involve process- and state-types that supervene on the current, internal, physical properties of organisms: “any differences between organisms which do not manifest themselves as differences in their current, internal, physical states ought to be ignored by a psychological theory.” If the Autonomy Principle can be justified, then the change in the status quo that Fodor also recommended would be justified: we ought not to cite contents as causes in our psychological generalizations.

Stich defends the Autonomy Principle as follows. Any person and her exact physical replica will behave in precisely the same ways in all circumstances; no one could tell the difference if Stich were replaced by his exact physical Twin. Of course, the behavior that will be precisely the same in these cases is *autonomously described* behavior. This is behavior under a description that is such that “if it applies to an organism in a given setting, then it would apply to any replica of the organism in that setting.” The reason that we want psychology to explain behavior under autonomous descriptions is that only under these descriptions will we be able to achieve the systematicity we seek in a scientific psychology: “It seems obvious that if we seek systematic generalizations to explain ... [our] behavior, we should not expect our generalizations to explain ... [our] behavior under *these* [non-autonomous] descriptions.”¹¹⁶ It would follow that the process- or state-types that should be used in psychological explanantia will be those that causally explain autonomously described behavior. But these process- or state-types must also obey the Autonomy Principle, for only states individuated by current, internal, physical properties can causally explain

¹¹⁶ Stich uses an analogy with the behavior of an industrial robot, and the original sentence is phrased in terms of explaining the robot’s behavior. But the analogy is to *our* behavior, so I have simply made that substitution in the brackets.

autonomously described behavior.¹¹⁷ Therefore, content properties ought not to be cited as causes in a scientific psychology.

There are several strands to this argument. One is the claim, addressed in detail in 2.3.2, that non-autonomous descriptions and the behaviors they pick out are “hybrids” of autonomous descriptions and the behaviors they pick out plus historical or other external facts and the relations and circumstances they pick out. I set this strand aside for now because even if true, the claim that non-autonomous behavior is a hybrid that can be decomposed into its internal and external parts is of no import unless we are given a reason for wanting to decompose it.

That motivation is provided by the claims about systematicity, which form the core of Stich’s Replacement Argument:

In evolving hybrid non-autonomous behavioral descriptions, common sense produces behavioral descriptions that are more fine grained than those that would be available if we restricted ourselves to autonomous descriptions. There is nothing unreasonable about this, since often enough our practical concerns demand some more fine grained description of behavior. But if I am right, then these practical concerns lead to a taxonomy of behavior which is ill suited to a systematic science aimed at explaining behavior. The thrust of the autonomy principle ... is that by building historical, contextual, and ideological features into mental state descriptions, folk psychology has taxonomized states too narrowly, drawing distinctions which are unnecessary and cumbersome when we are seeking a systematic causal explanation of behavior. By slicing the pie too finely, [wide individuation] impedes the formulation of those generalizations that ought to apply equally to an organism and its replica.¹¹⁸

The systematicity here is relative: autonomously described behavior (or mental states) is systematic *relative to* the requirements of causal explanation. Relative to these requirements, non-autonomously described behavior (or mental states) is not systematic.

¹¹⁷ Although this and the preceding sentence are not stated explicitly, they (or something like them) are required to complete Stich’s argument.

¹¹⁸ Stich, 169-70. For this reason, the “replacement argument” really ought to be called the argument from systematicity. But I think the label “replacement argument” is too entrenched to bother trying to change.

Thus, Stich's basic claim is that autonomous descriptions are systematic in ways relevant to causal explanation that non-autonomous ones are not, so a scientific psychology ought not to involve a taxonomy of non-autonomous behaviors and mental states.

My response, Part I. Stich provides only weak intuitive support for this claim about systematicity, in the form of (1) a few non-autonomous descriptions that do not seem amenable to playing a role in causal generalizations, and (2) a lack of specificity regarding the autonomous descriptions that allegedly do. Regarding (1), the fact that some behavior under non-autonomous descriptions is not systematic is no argument for the claim that they all fail to be. The success of our ordinary psychological explanatory framework provides *prima facie* evidence that there are at least some such behaviors (and explanatory states) that are systematic. So what Stich needs to provide is an argument that ties non-autonomy generally to the failure to be systematic in the relevant way – one that concludes that non-autonomous descriptions are not relevantly systematic in virtue of their non-autonomy. Similarly, regarding (2), Stich needs to provide an argument that ties autonomy to the ability to be systematic in the relevant way – one that concludes that autonomous descriptions are relevantly systematic in virtue of their autonomy.

What we get, however, is just a rough indication of some supposedly autonomous descriptions and their explanatory role. The autonomous description that Stich does provide – “performing a weld”, as opposed to the non-autonomous “performing a millionth weld” – is not autonomous, since welding presupposes all manner of external relations and objects. So this example does not even begin formulate a general argument.

However, without further guidance, we do know that what Stich and his replica have in common are their neurons and muscle fibers. Some of these same muscle fibers

will contract when someone raises Stich's arm or his replica's arm, and some of these same neurons and muscle fibers will fire or contract when Stich raises his arm or his replica raises *his* arm. So perhaps we can derive the general argument that Stich needs from these known considerations about what autonomous descriptions must be like. There are, I think, two possibilities available for this purpose.

The first is to interpret "autonomous description" such that it refers to some sort of muscle-fiber contraction descriptions.¹¹⁹ If we generalize from this interpretation, the sort of systematicity relevant to psychological causal explanation is systematicity under muscle-fiber contraction descriptions. But this means Stich's call for autonomous descriptions is a call for psychology to replicate the causal explanations of physiology. There may be arguments, such as those for eliminativism, that lead to the conclusion that we ought not to do psychology at all, but that is a change of topic (see 1.7.1). Our concern here was why wide contents shouldn't be used in psychological causal explanation, not why we shouldn't take any intentional claims seriously at any time.

Burge's response. Burge offers two detailed responses along the same lines of this first interpretation.¹²⁰ First, according to Burge, the only interpretation of "behavior" in which replicas behave the same is bodily motion (what I called muscle-contraction descriptions). But restricting psychology to explaining bodily motion would be a drastic, and unjustified, revision. For "many relevant specifications of behavior in psychology are intentional, or relational, or both." "Behavioral" evidence in psychology includes verbal

¹¹⁹ I am avoiding using the phrase "bodily motion" because that phrase is often used ambiguously to refer either to muscle contractions or to basic actions. I am disambiguating this usage in the text.

¹²⁰ Burge (1986: 10-12), Peacocke (1993: 207), Dretske (1988: 34-6) and Devitt (1996: 294-5) offer responses similar to Burge's. I will argue below that details about what intentional states *actually* explain do not address the main issue in Stich's argument, which is the connection he makes between relevant systematicity and autonomy.

behavior, such as questions asked of subjects and their replies; this will differ between replicas if externalism is true (and Stich accepts externalism). Also, many attributions of non-verbal behavior in psychology are non-autonomous: she picked up the apple, pointed to the square block, tracked the moving ball.

Second, psychology is not just the study of behavior even if “behavior” is construed as bodily motion. It includes theories of vision, memory, learning, linguistic understanding and other areas which involve explaining relations between subjects and their environments and help motivate externalistic individuation in psychology.

So, Burge argues, the demand for bodily motion as the proper subject of psychology entails a wholesale revision of the science in the direction of complete individualism, and this wholesale revision has not been supported. Burge does not claim that all of psychology requires externalistic individuation or that a new individualistic vocabulary and level of explanation – such as a computational level – cannot be added to it, but that at least some parts of it are not individualistic and that no good reason has been provided to eliminate them.

My response, Part II. There is also a second possibility for interpreting “autonomous description” in which the phrase refers to some sort of basic action descriptions. On this interpretation, the sort of systematicity relevant to psychological causal explanation is systematicity under basic action descriptions. This seems much more plausible than the muscle-contraction interpretation: Stich’s “calling a taxi” and his replica’s “calling a twin-taxi” (where these differ in the usual thought-experimental way) would count as the same behavior when redescribed as “moving one’s arm and hand back

and forth in the air.” The exact physical descriptions might differ even for just Stich on two occasions of taxi-calling, but that fine-grainedness won’t matter.

However, this basic action redescription (assuming it is one) requires that Stich and his replica have *intentions* that can share a description. In this case, the intention to move one’s arm just so is the common redescription of the intentions to call a taxi and to call a twin-taxi. If the redescribed intentions are narrowly individuated, these basic action descriptions satisfy the Autonomy Principle.¹²¹ But the problem this poses for Stich is that the systematicity of these autonomously described behaviors is due to the sameness in the contents of the intentions. It is not necessary that Stich and his twin replicate neural firing patterns or muscle contractions in order for them to have identical intentions and perform identical autonomous behaviors.¹²² To the contrary, we would lose the systematicity these descriptions provide if we abandoned them for (also) autonomous but highly fine-grained neural firing pattern and muscle fiber descriptions.

But if neural firing pattern and muscle fiber descriptions are *not necessary* to obtain the systematic descriptions needed for scientific psychological causal explanation - descriptions in which Stich and his replica behave the same -- then the Autonomy Principle is *unmotivated and unsupported*. This is so even if “autonomous descriptions” are neither muscle-fiber contraction nor basic action descriptions, but something else (computational descriptions, perhaps). For basic action descriptions with narrowly individuated intentions are just one type of autonomous description that are systematic

¹²¹ That is, I am supposing for the sake of argument that there is a narrow notion of intention (like any narrow notion of thought, individuated computationally or in narrow-content terms) such that this narrow state would cause bodily motion, making the behavior (on any view of behavior) narrow. The idea is to grant Stich a narrow notion of behavior that (plausibly) is *not* mere bodily motion but is not widely individuated. I will argue that the fact that it’s *narrow* (or *autonomous*) isn’t what makes it *systematic*.

¹²² Dretske (1988: 131-2, 145) remarks that the neural activity in the motor control system that brings about a given bodily movement is never precisely the same: “I move my arm each day, but at a neuronal level of description I probably never move it in exactly the same way.”

without their systematicity arising in virtue of the fine-grained physical states on which they supervene. Thus the fact that these behaviors are autonomous is *irrelevant* to their systematicity.¹²³ And if their autonomy is irrelevant to their systematicity, then we have no reason to think that the *failure* to be autonomous (or to supervene) is relevant to systematicity either. Non-autonomous descriptions may or may not be systematic, but whether they are or not is independent of their failure to supervene.

In short, to defend the Autonomy Principle, Stich needs to connect systematicity with autonomy; he needs an argument that shows that non-autonomous descriptions are not relevantly systematic *in virtue of their non-autonomy* or that autonomous descriptions are relevantly systematic *in virtue of their autonomy*. He has provided no such argument here, for we can obtain autonomous intentional descriptions that are systematic because they are intentional, not because they are autonomous. So autonomy is not necessary for systematicity. This leaves open the possibility that other intentional (but non-autonomous) descriptions are relevantly systematic. Moreover, there are plenty of autonomous descriptions that are not systematic – since no two real brains are exactly alike, any detailed autonomous description of one will differ from any other. So it not clear that autonomy is sufficient for systematicity either.

Of course, in all likelihood Stich implicitly links systematicity and autonomy by assuming that physically identical states will behave the same. This sameness of causal

¹²³ The response that these internal properties are systematic because they can be *reduced* to the properties on which they supervene, even if true, would change the subject by strengthening the Autonomy Principle such that it makes a reductionist, rather than non-reductionist, claim. My argument assumes supervenience is supervenience, not reduction. However, this response points to an important background issue in the Arguments from Narrow Explananda: whether reduction is required for causal efficacy. These worries about the mental-physical property relation help motivate the account I give in chapter 4.

power in turn would ground causal explanatory systematicity.¹²⁴ The cases of chiral molecules should raise doubts about that assumption; and an explicit argument involving that assumption will be discussed in Ch. 4, when I respond to Fodor's Non-Locality argument. The point now is just that Stich has not argued for that assumption, even though it may be doing some work in his Replacement Argument.

Summary. Stich justifies the Autonomy Principle as a criterion of adequacy for a scientific psychology by arguing that properties that do not satisfy the principle will not be systematic in the way needed for causal explanation. The Autonomy Principle covers properties that individuate mental states and those that individuate behavior. Stich does not argue directly against the systematicity of wide contents; instead, he argues against the systematicity of wide behavior (or wide behavioral descriptions). This indirect strategy logically requires a second argument to the conclusion that *only* autonomously described events ("mental" or otherwise) can be causes of autonomously described behavior.¹²⁵ No argument of this sort is offered. So the overall argument to the conclusion that wide contents ought not to be cited as causes in serious science remains incomplete whatever the fate of Stich's argument about behavior.

That said, his argument about behavior fails. I argued that there can be autonomous properties that satisfy the systematicity requirement because they are intentional, not because they satisfy the Autonomy Principle. It is not enough to argue that autonomous descriptions are relevantly systematic; the autonomous descriptions must be relevantly systematic *because* they are autonomous. Without making this connection, the Autonomy Principle appears irrelevant to the aim of systematic causal

¹²⁴ Burge (1986: 13-20) also notes as a "kindred argument" the claim that "the determinants of behavior supervene on the brain." I will discuss this in the next chapter.

¹²⁵ This issue will come up again in my discussion of Kim (in 2.3.2) below.

explanation. So we have no reason to accept it as a criterion of adequacy for causal explanation. So Stich gives us no reason to deny that contents are causes.

In addition, Burge argues that Stich's recommendation that psychology limit itself to autonomously described behavior entails a drastic revision of current psychological practice. Psychology considers non-autonomously described verbal and other behavior as its explananda, and in any event it also is concerned with explaining a wide range of cognitive and perceptual abilities, not just behavior however construed. The weakness of this sort of response is that it does not address Stich's main claim that only autonomously described events are relevant to causal explanation.¹²⁶ Pointing out that psychology explains relationally, intentionally, or non-autonomously described behavior is meaningless unless it is shown that these properties are relevant to causal explanation. So Burge's response is at best incomplete.

2.3.2. Kim's Causal Core Argument

Argument. Kim's Explanatory Thesis plays a role similar to that of Stich's Autonomy Principle and Fodor's Formality Condition in a prescriptive argument to the conclusion that we ought not to cite (or consider) intentional states as genuine causes.¹²⁷ The Explanatory Thesis states that "internal psychological states are the only psychological states that psychological theory needs to invoke in explaining human behavior – the only states needed for psychology." If supported, this would justify an entirely individualistic psychology, not just in the explanation of behavior, but also in every other branch of psychology.

¹²⁶ Stich (1983: 169) never denies that wide properties can serve certain explanatory purposes, such as those involving our "practical concerns"; he denies that they are useful in causal explanation specifically.

¹²⁷ Kim (1982/1993: 183-191).

Kim defines internal states or properties as non-relational in both time and space: such states do not imply anything about the past or future or about anything existing other than the organism that has them. (See 1.5.1) Kim's strategy is to argue:

that within each noninternal psychological state that enters into the explanation of some action or behavior we can locate an 'internal core state' which can assume the causal-explanatory role of the noninternal state; we would in fact argue that this internal core is the causal and explanatory core of the noninternal state. It is in virtue of this core that the noninternal state has the psychological explanatory role that it has.¹²⁸

There are two claims here that Kim must defend: (1) that each noninternal psychological state contains at least one state that is both *internal* and *psychological*; and (2) that "the causal-explanatory role of any noninternal psychological state can be filled by some internal psychological state." These claims together comprise the Explanatory Thesis.¹²⁹

Along with the thesis, Kim also appears to assume the traditional connection between causation and causal explanation (see 1.4.3), in which a genuine causal explanation is one that is ontologically grounded in a singular causal relation between events in virtue of properties they have that (in the best of cases) are explicitly referred to in the explanation. For on Kim's view, internal states are the only states needed to *explain* behavior because they are the only states that are "causally productive" of behavior; they are "the causal as well as the explanatory" core of the wide states. Kim also awards the usual physicalist booby-prize to intentional psychological explanations, which are granted second-class status as "rationalizing explanations" since they aren't

¹²⁸ *Ibid.*, 189-90.

¹²⁹ The Explanatory Thesis also claims that internal states are *all* that is needed in psychology, but this pan-individualist claim is not defended by Kim, and in any case Burge's response to Stich would apply here as well. However, for the reasons noted in the previous section, this response would not address the core issues being raised by Kim, just as it did not address the core issue raised by Stich.

genuinely causal.¹³⁰ Given his assumption about the relation between causation and causal explanation, the reason these explanations aren't genuinely causal is because wide intentional properties aren't genuine causal powers. And the reason why wide intentional properties aren't genuine causal powers is because of the Causal Core argument, which has the Explanatory Thesis as its essential premise.

The Explanatory Thesis could use some clarification. Kim's remarks regarding (2) are ambiguous, for the claim that the causal-explanatory role of an external state can be filled or assumed by an internal state is weaker than the claim that the internal state is that in virtue of which the external state has the causal-explanatory role it has. The former interpretation leaves open the possibility that the external state can play a causal-explanatory role too; it does not entail that relational properties are causally inert. The latter interpretation entails that the external state can play no causal-explanatory role but for that played by its causal core; it entails that its relational properties are causally inert.

There is also an ambiguity in Kim's remarks regarding (1), in particular regarding the relation between the internal and external states. On the one hand, he claims that the internal state has the causal powers of the "corresponding" noninternal state; this might imply that the internal state is a distinct particular that replaces the external state. On the other hand, he writes of a memory being "stripped down to an internal psychological state", as if we peel away its relational properties without affecting the identity of the state. It remains the same state. It is also open what this "stripping down" amounts to: for example, the properties that characterize the internal state may be abstracted from those of the external state (as 'animal' might be abstracted from 'dog'), or the internal state

¹³⁰ See Kim (1984). This standard response was discussed in 1.4.3.

may be a proper part of the external state (as bodily motion may be a constituent of intentional action).¹³¹

Despite this unclarity, I think Kim's intent in (1) is to refer to a single state to which some sort of paring operation is applied in order to obtain a remainder that is purified of causally irrelevant properties (or causally irrelevant components of complex properties); and that his claim in (2) is the stronger claim which eliminates the possibility that external properties are causally relevant. If this is right, then intuitively the Explanatory Thesis holds that a wide mental state can be pared of its relational properties like so much excess fat, leaving behind the *same* state, but trim and fit for causal explanation, because now it has only those properties that may be genuine causal powers.

This "causal core" idea, as applied to both intentional states and behavior, is an extremely popular interpretation of extrinsic mental states. Block defends narrow content as "a nonrelational aspect of propositional attitude content, the aspect 'inside the head,' that corresponds to content the way moving the fist corresponds to hitting."¹³² Stich calls non-autonomous behaviors "logical or conceptual hybrids" which are "resolvable into an autonomous component and a potpourri of other factors which explain why the autonomously described event *counts* as satisfying the non-autonomous description."¹³³ Pettit examines the explanatory role of intentional states on the assumption that they can be "decomposed" into a psychological part and a non-psychological environmental component in such a way that the external state is related to an intrinsic "decompositional

¹³¹ Another popular term for the operation is "factoring", as in (e.g.) Field (1988: 103), where we "factor out" Twin-Earth (relational-property) differences between (e.g.) gold and twin-gold. In Field's case at least, the factoring out appears to be abstraction, e.g. from "gold" and "twin-gold" to "shiny metal stuff."

¹³² Block (1986: 626).

¹³³ Stich (1983: 168-9).

alternative” as fragility is related to molecular structure.¹³⁴ Schiffer offers an argument (on behalf of an imaginary physicalist) that is effectively identical to Kim’s: “[O]nce we have explained bodily movement we have gone, it would seem, as far as psychology can legitimately take us ... [I]t does indeed appear that the only legitimate task for psychology as regards the explanation of behavior resides in the explanation just of bodily movement, and for that [narrow] conceptual role psychology suffices.”¹³⁵ Thus, Kim’s “causal core” idea is representative of a host of similar views, and his argument regarding the causal inefficacy of wide contents would apply to any of them.

Kim’s argument for the Explanatory Thesis (or the “causal core” idea in general) rests on a series of examples. Regarding (1), he claims, we can imagine some internal core of every typical wide intentional state: a generic “seeming to ____” state. Memory is noninternal because it implies something about the past and the existence of something other than the person remembering. But “we may assume there obtains in him some internal state which is just like a genuine case of remembering except for one of these conditions failing to obtain.” Similarly, knowing is a state that can be decomposed into an internal state of believing plus a relation to the world in virtue of which the belief is true.¹³⁶

The argument for (2) also relies on examples. Knowing (believing truly) is a noninternal state in that if I know *that the burner will go on when I turn the knob*, it

¹³⁴ Pettit (1986). Pettit argues that the wide explanations are indispensable for ‘normalizing’ (reason-giving) explanations, but they remain dispensable for ‘regularizing’ explanations – explanations that “present the event explained as an unsurprising moment in the causal unfolding of the natural world.” This is an example of the ‘proleptic’ move noted in 1.4.3., one which deprives relational properties of the genuine causal efficacy that I aim to defend.

¹³⁵ Schiffer (1981: 217-9). He responds to this argument that wide psychological explanations are needed because we are ignorant of all the environmental facts that we’d need to add to the conceptual role explanation to predict ordinary behavior. At best, this defense means the psychological explanations are only pragmatically, not causally, required; it is not the defense of intentional causation that I hope for.

¹³⁶ Pettit (1986: 19) also uses the popular example of knowing as a decomposable state.

follows that the knob and the burner exist. But knowledge is “sufficient but not necessary to construct an action explanation ... In fact it is only the element of belief in knowing that is causally productive of the action”, where the action that Kim has in mind is “the bodily action of turning the knob”. This is a “basic action” for Kim, and “only basic actions ... need to be explained by psychological theory.” While my action of *turning on the burner* surely depends on whether my belief is true, “it is not part of the object of *psychological* explanation to explain why the burner went on; all it needs to explain is why I turned the knob.” Similarly, the internal state of seeming to remember “can do all of the explanatory work done by remembering.” If I act in a certain way because I remember something, “my replica will act in the same way because of his seeming to remember the same thing. Whether or not his seeming remembrance is a genuine case of remembering will not affect his behavior.”

Kim’s defense of the claim that the causal cores are *psychological* is as follows. When we are constructing a psychological explanation of behavior, we are seeking a psychological causal antecedent of the behavior. “Considerations of causal contiguity and continuity lead to the belief that the proximate cause of the behavior must be located within the organism emitting the behavior – that is, there must be a proximate causal explanation of that behavior in terms of an *internal* state of the organism.” This internal state is justifiably considered psychological because “if it has all the causal powers of the corresponding noninternal state in the production of behavior, then there seems to be no reason not to think of it as psychological as well.”

My response. As a defense of the claim that the causal cores are psychological, the argument of the previous paragraph is harmless; if you have an internal state that does

what the external state does, then there's no reason to deny it psychological status. The existence of internal causally efficacious content properties is no threat to the externalist's position.¹³⁷ What is significant in the argument is that it is the closest Kim gets to facing the obvious problem with the entire causal core idea: the proper application of the chosen method of obtaining the causal core – abstraction, subtraction, factoring, decomposition, deanchoring¹³⁸ -- *presupposes* some criterion of genuine causal efficacy in virtue of which the operation reliably yields a purely causal core. This unstated criterion must *already* rule out relational properties as causally efficacious. For by applying this criterion to externalistic states, we eliminate just those properties that are causally inert – *et voilà*, the result is an internal, causal core. But of course a *different* criterion would also eliminate all causally inert (by its lights) properties yet yield a causal core that isn't *internal*. In short, the causal core argument is superfluous as a means to the conclusion that we don't need relational properties for causal explanation. Whatever defense Kim might have given for his criterion is sufficient to obtain this result.

The significance of this argument, then, lies in Kim's reference to the need for internal states due to "considerations of causal contiguity and continuity." Presumably these "considerations" support his unstated criterion of genuine causal efficacy that is needed to factor out the causal core. They would thus play a key role in an argument for the crucial claim that the internal state is a *causal* core (from which it might *follow*, if the causal roles are the same, that the state is *psychological*). That argument might go

¹³⁷To make my point clear: the issue of whether the internal states are psychological (that is, whether they count as a kind of *contentful* state) may play an important role in the literature on narrow content, but it is irrelevant here. The relevant issue is whether the internal core has all the causally relevant properties of the original external state; this claim entails that the relational properties are causally irrelevant.

¹³⁸This last is Fodor's (1987: 49-51) suggestion, where an expression is anchored iff it has a determinate semantic value. He warns against thinking of de-anchored (i.e. narrow) content as content obtained through subtraction, and uses instead the analogy to the way that the content of demonstratives or indexicals ("I", "that", etc.) is analyzed as a function determined by are contextually determined.

something like this: Causally efficacious properties satisfy a criterion of causal efficacy based on causal contiguity and continuity; relational properties violate this criterion; so they are not causally efficacious. Kim only provides a hint of this argument, but it is found explicitly in the Arguments from Non-Locality presented in Ch. 3.

I should add that none of Kim's explicit examples provide support for a criterion of causal efficacy that rules out relational properties. If externalism is true, belief is an external state whether it is true or not; as Burge's arguments for externalism (see 1.2.3) show, we credit the Patient with a *false* belief *about arthritis*. The fact that a belief is true has nothing to do with its being a relational state. Seeming to remember also need not be internal. A painting may only seem to be a Rembrandt, but it is still a painting with all the externalistic relations that being an artwork requires. Similarly, a seeming to remember may be a false memory, but its being false has nothing to do with its being relational. At best, Kim's examples provide evidence that modes of presentation of some sort play an indispensable role in psychological explanation. But since that is compatible with a referentialist, externalist theory of content (such as Devitt's), it cannot be an argument that beliefs are internal.¹³⁹

The same result occurs with Kim's examples regarding the proper explanandum of psychological causal explanation, even assuming that this explanandum is limited to "behavior". It is essential to the 'causal core' argument that the explanandum not involve an action under a relational description, for (as Kim is aware) if the explanandum involves some notion of success – such as in *turning on the burner* – then mere belief

¹³⁹Devitt (1996: 136-244) argues that while the different sorts of contents we ascribe are all wide – that is, they are determined directly or indirectly by noninferential causal links to the world -- opaquely ascribed contents "are the ones that ultimately explain intentional behavior." (p.185) In contrast, transparent ascriptions (p. 150-154) are designed to give us information about the world.

will *not* suffice to explain my action; even if we assume belief is internal, this internal state will *not* replicate the causal-explanatory role of the noninternal state of which it is the 'core' and (2) will be false. But Kim's example of a non-relational behavior -- *turning the knob* -- is just as relational as *turning on the burner*. I cannot perform this action unless I truly believe that there is a knob to turn; even worse, no purely internal state or combination thereof can explain my *turning the knob*.

If we grant Kim what he perhaps intended -- that my *moving my hand just so* is the action that it is the job of psychological causal explanation to explain -- this basic action is still an action, not a bodily motion. If the intention that is its cause is internally individuated, we may grant that this basic action is internal. But now the problem facing Kim is not unlike that which arose for Stich. Since Kim is using an indirect strategy -- arguing for narrow states in part by arguing for narrow behavior -- he must argue that only an internally individuated state can cause an internally individuated behavior. *Prima facie* this seems false: my (noninternal) belief that a tiger is charging at me can cause an adrenalin rush or a dry mouth. That is, even if narrow states suffice to causally explain narrow behavior, narrow states are *not necessary* to explain narrow behavior.¹⁴⁰

What Kim might say in response is that only the narrow cores of wide states -- such as my narrow-tiger-belief -- *really* causally explain narrow behavior; the wide states only *seem* causally sufficient because they contain internal cores that really *are* causally sufficient. But this just gets us back to the first part of my response. That the internal states are *genuinely causal* cores simply follows from applying Kim's unstated criterion of genuine causal efficacy to pare away the extrinsic properties (or the extrinsic

¹⁴⁰ The problem for Stich was that we could identify systematic, autonomous behavioral descriptions in which their autonomy was not necessary for their systematicity. Kim's problem is internal causal cores don't seem necessary (*prima facie*) to cause narrow behavior.

components of complex properties). Kim might as well just point to the Non-Locality arguments of the next chapter.

Peacocke's response. Peacocke provides an example of a common response to the "causal core" line of argument, although he is not responding to Kim in particular.¹⁴¹ The response is that wide explanations involve a different sort of explanandum; they explain relationally described behavior. In Peacocke's terms, the worry that wide states are causally irrelevant:

is based on a misapprehension about what is explained by states with mental content. It is true that brain states and efferent connections will be enough to explain any given bodily movement. But what is distinctively explained by a set of externalist states are relational facts which go beyond mere bodily movements. ... [T]here are distinctive counterfactuals linking relational causes and relational effects with the presence of externalist states. If the subject had not seen the person as being in the garden, he would not have pointed towards the garden. It is counterfactuals like that which support the explanatory importance of externalist states.¹⁴²

Thus, the fact that Kim argues that narrow states explain narrow behavior does not threaten the claim that external states explain external behavior (or more precisely, that relational properties of mental states explain relational properties of behavior).

Peacocke's suggestion that there are distinct explananda, each of them causally explained by different states, is accompanied by the claim that the explanatory importance of externalist explanation is supported by counterfactuals. The two claims are closely linked. Wide states are explaining something different from the usual causal chains we are familiar with, and what counts as a causal explanation is different as well. This latter claim, however, is critical to assessing the strength of Peacocke's response. For Kim can cheerfully agree that the wide states have their own explanatory importance

¹⁴¹ Peacocke (1993). Similar responses are given by Burge (1995) and Baker (1995).

¹⁴² Peacocke (1993: 207).

but deny that psychological explanations are genuinely causal. So for a complete response Peacocke must also defend the idea that the truth of the appropriate counterfactuals is sufficient to justify the causal status of these explanations.

Notably, this dialectic is repeated between Fodor and Burge in Chapter 3. Kim and Fodor hold the traditional view in which explanations are causal if they are grounded in singular causal relations between events, while Burge and Peacocke hold the counterfactual view, in which explanations are causal if they are supported by the relevant counterfactuals. There is little or no explicit discussion in either case of this fundamental difference. One major reason for this, I think, is that Kim and others who make “causal core” arguments fail to make explicit, and argue for, the criteria for genuine causal efficacy that are presupposed in the argument. (Fodor’s Non-Locality argument makes these criteria explicit.) But by the same token, those who respond to the explicit aspects of the ‘causal core’ arguments are mistaken if they think that their opponents accept the same criteria for genuine causal explanation that they do.

This is why I find the sort of response to Kim exemplified by Peacocke weak. As a defense of psychological causal explanation, Peacocke relies on a view in which what makes an explanation causal is not that it is grounded in a singular causal relation between events, but that it is supported by appropriate counterfactuals. That this is sufficient for genuine causal efficacy needs more argument than it is given. (What little Burge has to say about this is provided in Ch. 3.) Moreover, I argued in 1.4.3. that a strong defense of genuine causal efficacy for relational properties should, if possible, follow the standard account of the relation between causation and causal explanation. A response that points out that the explananda are distinct risks putting wide intentional

causes in the analogous position as segregated schools: separate and unequal. Even worse, this sort of response ignores the fundamental objection – expressed in the Non-Locality arguments – that there is something about relational properties that makes them unfit for causation. It may turn out that there is nothing wrong with them; I argue this in Chs. 3 and 4. But so far those types of arguments have not yet been given.

Summary. Kim claims that all we need to explain in psychology is internally described behavior, and he argues that this can be causally explained by internal psychological states that can be carved out of ordinary external states. Even if we forget about all other possible psychological explananda – vision, memory, and so forth -- for the sake of argument, Kim's argument does not adequately support his conclusion. His argument presupposes that relational properties violate an undefended principle about what is necessary for genuine causal efficacy. This principle is needed in order to carve out the internal causal cores. Without an explicit defense of that principle, Kim's argument presupposes what it must prove: that relational properties are not causal. The unstated argument supporting his criterion for carving out causal cores is doing the work.

Peacocke argues that relational states play a legitimate and distinctive explanatory role in the explanation of relationally described behavior. This response depends on a view of what makes an explanation genuinely causal that Kim does not share and so which cannot be taken for granted. So even if Peacocke is correct that relational states explain relationally described behavior, the question remains whether these explanations are truly causal. Until that question is addressed, the sort of response exemplified by Peacocke will always be subject to some version of the rejoinder that relational properties are causally *relevant* but not *efficacious* (see 1.4.3). But causal relevance is a much

weaker notion of causal involvement than the robust sense of causal efficacy that the defender of intentional causes ought to aspire to.

2.4. FODOR'S CROSS-CONTEXT AND MODAL ARGUMENTS

Two further arguments against the causal efficacy of wide content are provided by Fodor: the Cross-Context argument and the Modal argument. These are (logically) preliminary to his (1987) Non-Locality argument, which I consider both his main objection and the main objection overall to causally efficacious wide contents.

In the (1987) Cross-Context argument, he argues that we would judge that the effects of distinct wide contents in the same contexts would be the same. This would violate what Fodor considers a general principle in science: that no property counts taxonomically – that is, classifies entities for the purposes of causal explanation – unless it affects or makes a difference to causal powers. If the Cross-Context test is right, then wide content differences would not make a difference to causal powers, and so would not count as causal powers in science. In the (1991) Modal Argument, Fodor argues that the differences between wide contents and their respective behavioral effects are conceptually connected, and therefore that these differences do not count as causal-power differences. Again, the conclusion is that a taxonomy by wide content is not a causal-power taxonomy. I should note that both arguments are intended to support a notion of narrow content, and that Fodor later repudiates narrow content.¹⁴³

¹⁴³ See Fodor (1994: 28-29), where he argues that narrow content is “superfluous.” However, he rejects his Modal argument because he rejects Twin cases as being only conceptually, but not nomologically, possible. If there *were* (lots of) Twin cases, then empirical psychology that taxonomizes using wide contents would not sort things by their causal powers. But since there *aren't* lots of Twin cases, a wide content psychology is perfectly acceptable; it will capture the nomologically possible generalizations it needs to. I discuss his (1994) solution further in chapter 3.

2.4.1. Fodor's Cross-Context Argument

Argument. Fodor begins his case against admitting some relational properties as “causal properties” by arguing that causal powers are individuated across contexts, not within contexts. He then claims that externalistic contents do not pass this cross-context test. To individuate across contexts is to make judgments of sameness and difference while keeping contexts constant. If we judge that our causal powers would have the same effects in all the same contexts, then our causal powers are the same. Causal powers are assessed *ceteris paribus*, where what's equal includes equal contexts. In slightly more formal terms: P, which means c in @ and causes e in @, *would have caused* e in the closest nomologically possible W in which P doesn't mean c.¹⁴⁴ If the cross-context test shows that causal powers are the same even when wide contents differ, then wide contents differences are causally irrelevant. What matters causally would be some property that both states have in common – in Fodor's view, their narrow content.

Fodor's reason that the identity of causal powers “has to be” assessed across contexts begins with an intuitively compelling example. Surely my power to lift a chair and your power to lift a chair are the same if we can lift it in all the same contexts, and differ if I can lift it in at least one context in which you cannot. But if you can lift a chair in a context in which it is not nailed to the floor, and I can't lift it in a context in which it is nailed to the floor, “that eventuality would give your biceps nothing to crow about”: it would indicate nothing about whether our causal powers are the same or differ. For we

¹⁴⁴ Fodor (1987: 34), adjusted to the case of content. The original version is presented in terms of his famous H-particle example, in which a particle is an H-particle just in case the dime in Fodor's pocket is heads up (*mutatis mutandis* for T-particles). Note how the property of being an H-particle is a mere Cambridge property (see 1.5.2.), since whether a particle's being an H-particle does not depend at all on how things are with the particle. The example is thus a rhetorical device in which we are invited to assimilate relational properties in general to mere Cambridge properties. However, Fodor abandons this line when he accepts that some properties (such as being a planet and being a meteor) are relational and causally efficacious. I will discuss these cases in the next chapter.

would be assessing my causal powers in one context and assessing yours in another, and so we would not be making a fair, across-context comparison of our causal powers.

Similarly, Fodor argues, it is true that when I say “water” I get water and when my Twin says “water” she gets twater. But these effects of our causal powers only differ because they occur in different contexts, and we cannot conclude anything about the sameness or difference of our causal powers based on differences in effects that occur in different contexts. Such judgments:

... are irrelevant to the question of the 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. For our utterances (/thoughts) to have the same causal powers, both of these counterfactuals have to be true. But both of these counterfactuals *are* true, since (for example) if I had said “Bring water!” on Twin-Earth, it’s XYZ that my interlocutors would have brought; and if he had said “Bring water!” here, his interlocutors would have brought him H₂O.¹⁴⁵

Thus, Fodor’s water –beliefs and –utterances are the *same* causal powers as his Twin’s twater –beliefs and –utterances. While their contents differ due to relational differences, they are identical *as causal powers*.

My (and Burge’s) response. The case Fodor presents does not support his claim that relational intentional properties are not causally relevant. This can be seen if we examine separately the two quite different judgments of individuation that can be made in the case he cites: regarding the sameness of our utterances qua effects of our thoughts, and regarding the sameness in others’ behavior qua effects of our utterances.

In the first case, the cross-context test does not rule out the causal relevance of the relational properties of our thoughts, since our utterances (qua effects of our thoughts)

¹⁴⁵ Fodor (1987: 35).

will differ across contexts. Given my water thoughts, I will ask for water on Earth *and* Twin-earth, and given her water-thoughts, my Twin will ask for twater on Earth *and* Twin-earth. We would not conclude that the differences in our contents is not a difference in causal powers, for both counterfactuals would be false. So *pace* Fodor, in this case wide contents seem to be bona fide, distinct causal powers.

The second case is trickier. In cases of nonverbal behavior, Fodor argues that our causal powers will be the same because in a watery context, I get water when she gets water, and in a twatery context I get twater when she gets twater. Fodor concludes: same effects across contexts, same causal powers. But the fact that others bring us both water on Earth and twater on Twin-earth in response to our utterances is surely a function of what's available for them to bring us on Earth and Twin-earth. This seems irrelevant to assessing *our* causal powers – what I can do and what my Twin can do. As Davidson – or even Kim – might put it, all I ever do is try; whether the world cooperates, and how it does, is not up to me.¹⁴⁶ The same response would hold even if we eliminate other agents. When my Twin and I get the water or twater ourselves, we both get water on Earth and twater on Twin-earth, but only because in each context that's all there is for us to get.

The problem – as Burge also points out¹⁴⁷ – is that value of the cross-context test as a way of determining sameness or difference of causal powers depends entirely on the nature of the contexts. Some contexts don't provide any opportunity to demonstrate any difference in causal powers, should there be one. My Twin and I both get water on Earth and both get twater on Twin-Earth. But so what? Given these contexts, the fact that we'd judge the appropriate counterfactuals as true when comparing causal powers across them

¹⁴⁶ Davidson (1971); reprinted in his (1980: 59). Kim's distinction between turning the knob vs. turning on the burner is an attempt to draw this distinction.

¹⁴⁷ Burge (1989: 311-313).

doesn't mean we have the same causal powers. We might or might not. Suppose I can benchpress 500 pounds and you can benchpress 50 pounds. Suppose we compare our causal powers across these contexts: one which only has weights of 51 pounds or more, and one which only has weights of no more than 50 pounds. The counterfactuals above are false, and Fodor would conclude that we differ in causal powers. But suppose we compare our causal powers across these contexts: one in which there are no weights of more than 50 pounds and another in which there are ordinary chairs nailed to the floor. The counterfactuals would be true, and Fodor would conclude that we have the same causal powers.

In short, the only contexts that matter for judging sameness or difference of causal powers can be those in which a difference can show up if there is one. So what the cross-context test should require is that if we have the same effects in all the same *relevant* contexts, then our causal powers are the same. The *relevant* contexts are those where a difference in causal powers could show up if there is one. So the validity of the cross-context test as a way to rule out wide contents comes down to a debate over which contexts are relevant.

Now if, as Burge notes, the contexts include ones where both water and twater are found and where there are people who have the concepts of water and twater, in those contexts when I ask for water I get water and when my Twin asks for twater she gets twater. Since our effects differ across these contexts and the counterfactuals turn out false, we can reasonably infer that our wide contents are distinct causal powers. Burge adds that speaker understanding is inessential; for *any* causal-historical property, a device could in principle be built that detects whether individuals have that causal history – past

contact with water, for example. The device could bring water to those individuals who exhibit the causal-historical property it detects and do nothing otherwise. This would be a difference in effects from which we could infer a difference in causal powers.

More generally, we must already possess *some* idea of what a science's subject matter is – what it is trying to causally explain – in order to begin to identify the set of contexts relevant for determining causal power differences:

[I]nterpreting the test in such a way as to make it individuate those causal powers that are relevant to scientific typology requires that one already know what counts as an admissible kind. ... [I]t is hardly obvious that a context in which a speaker's utterances are understood is an inappropriate context for checking causal powers relevant to psychological explanation.¹⁴⁸

In the present case, it is question-begging to assume that contexts in which causal-historical properties *could* make a difference are irrelevant. But in those contexts, the differences in the effects of water- and twater- beliefs are apparent.

Fodor's own rejoinder, and responses. Fodor imagines a different response to his argument: that the Twin behaviors differ across contexts when intentionally described, even if they do not when they are nonintentionally described. With this response Fodor shifts his focus from the case of what my utterances cause *others* to do to the case of my thoughts causing what *I* do. As noted, the cross-context test does not rule out wide contents as causal powers in the case of verbal behavior: my thoughts cause water-utterances and my Twin's thoughts cause twater-utterances, and these effects differ across contexts. Fodor avoids utterances and introduces a new nonverbal example: my desire to buy brisket causes my brisket-buying behavior and my Twin's desire to buy

¹⁴⁸ Ibid., 312. He adds that the test "is insensitive to differences among the typologies of different sciences. So it is insensitive to the different ways in which sciences recognize causal powers." His example of a heart, which is individuated in physiology by evolved function, is supposed to show that while the causal powers of the heart and a Twin as typed by physics may be the same, they do *not* have the same causal powers as typed by physiology (which takes causal-historical relations into account).

brisket₂ (which is limited to beef) causes her brisket₂-buying behavior. “Brisket-buying behavior” could be behavior which results in the buying of brisket, which can’t occur in a world without quadrupeds, or it could be behavior caused by a brisket-intention, whether or not there are quadrupeds to be slaughtered. Fodor does not distinguish between these.¹⁴⁹ Either way, since these are different effects (intentionally described) across contexts, the response goes, we can infer to different causes.

Of Fodor’s two rejoinders to this imagined response, one launches his Argument from Non-Locality discussed in Ch. 3. He grants that we can finesse the Cross-Context argument by individuating effects (such as brisket-buying behavior) as different, just in order to ensure that the causes differ as well. But, he argues, wide contents would still not be causes because they would violate the “locality” of causation.

In the other rejoinder, he claims that even if we did individuate behaviors differently when they have different intentional descriptions, and from this infer that the causal powers of our desires and beliefs differ, we would have to infer that the causal powers of my brain states and my Twin’s differ, which is “grotesque”. For example, suppose I am in a brain state that yields “the form of words” ‘Bring water’ and she is in a brain state like mine that yields the same “form of words”. If these effects are counted as distinct, then by parallelism of argument our brain states would have to be counted as distinct as well.

I’ll argue in Ch. 3 that this possibility is not at all “grotesque” – brain states may very well be widely individuated. Burge, however, replies that there is no parallelism of argument: “psychology does not and need not individuate mental states in parallel with

¹⁴⁹ The difference plays a role in his (1991); see the next section.

brain states.”¹⁵⁰ Each science determines its own typology depending the sorts of causal explanations it seeks. There is no reason to think that psychology and neurophysiology must individuate their causal powers in tandem. I’ll discuss Burge’s reply further in Ch. 3, since it is part of his response to Fodor’s Non-Locality argument.

Summary. Fodor suggests a test for causal power individuation which, plausibly enough, rests on the intuition that if effects are the same in the same contexts, we can reasonably if defeasibly conclude that the causes are the same. I argued, with Burge, that until we have determined which contexts are relevant – that is, which contexts will allow for a difference in causal powers to show up if there is one – then the test is useless as a means for inferring sameness of causes from sameness of effects. Burge adds that if the chosen contexts are ones in which differences between causal-historical properties will not show up, the test is question-begging. In general, we must already have some idea of what kinds of phenomena a science is seeking to explain before we can determine the relevant contexts for testing identity of causal powers. That is, we can only legitimately judge sameness and difference of causes in relevant contexts, but we need to have already judged sameness and difference of effects to determine which contexts are relevant.

Fodor replies that if we individuate effects under intentional descriptions and from this infer to different causal powers, we would have to say that brain states also differ in causal powers even when they are molecularly identical. Burge responds that psychology and brain physiology need not individuate their causal powers in tandem, while I will defend the idea that brain states may be individuated widely. Both responses will be discussed in greater detail in Ch. 3.

¹⁵⁰ Burge (1989: 308). He also adds (fn. 4, p. 319) that brain states, as entities with evolved functions, are also individuated in relation to a normal environment, and hence do not supervene on atomic make-up.

2.4.2. Fodor's Modal Argument

Argument. Fodor (1991) takes another stab at trying to rule out wide intentional properties as causal powers and, by this indirect means, argue for a notion of narrow content. He begins by claiming, as a general principle, that the difference between one cause property (CP1) and another (CP2) is *responsible for* a difference between one effect property (EP1) and another (EP2) in that if a cause with CP1 had had CP2 instead, then its effect would have had EP2 rather than EP1; and if a cause with CP2 had had CP1 instead, then its effect would have had EP1 rather than EP2. This responsibility for a difference in effects counts as *causal* responsibility when a further condition, Condition C, is satisfied: "Only when it is not a conceptual truth that causes which differ in that one has CP1 where the other has CP2 have effects that differ in that one has EP1 where the other has EP2." Condition C is a justified condition on a property's being a causal power because, "as Uncle Hume taught us", causation is a contingent relation.

Fodor then argues that wide contents do not satisfy Condition C. The differences between their effects are conceptually connected to the differences in the wide contents:

For the difference between being CP1 and being CP2 to be a difference of causal powers, it must at least be that the effects of being CP1 differ from the effects of being CP2. But, I claim, it is further required that this difference between the effects be nonconceptually related to the difference between the causes. Broad-content differences, per se, do not satisfy this condition, however. There are differences between my behavior and my twin's which are due, in the first instance, to the difference between the intentional contents of our thoughts, and, in the second instance, to my being connected to water in a way that he is not. But these differences among the effects are conceptually related to the differences between the causes; it is conceptually necessary that being connected to water rather than twater leads to water thinking rather than twater thinking; and it is again conceptually necessary that water thinking leads to water behaving and twater thinking does not. (1991: 24)

Fodor does not claim that wide contents are not causal powers. They are. The problem is that wide content differences do not count as differences in causal powers, because the differences in their effects (as shown in Twin cases) are conceptually connected to the differences in the causes. This means a wide-content taxonomy is not a taxonomy by causal powers, although it may serve non-causal-explanatory purposes. A psychological causal taxonomy should classify mental states narrowly.

The argument is related to the “logical connection” argument in action theory. That argument, attributed to Ryle and G.E.M. Anscombe among others, claimed that there could not be a causal theory of action because there was a conceptual connection between the intention to do A and the intentional doing of A. Davidson replied that while we can give descriptions of causes and effects under which they conceptually connected, it does not follow that the relata are not causally related. For they may stand in a causal relation under other descriptions that are contingently related.¹⁵¹

Similarly, while we can give descriptions of mental causes and behavioral effects in terms of wide contents and intentionally-described behavior, if Fodor is right this does not show that these events stand in causal relations under these descriptions. His reasoning appears to be that (1) since water- and twater-behaviors differ *only* in their intentional descriptions, and (2) since the difference between these intentional descriptions is conceptually connected to the difference between their respective mental causes, then (3) we have reason to think that the difference between water-thoughts and twater-thoughts is not a difference in causal powers.

¹⁵¹ See Davidson (1963/1980: 3-19); he cites many sources of the “logical connection” argument in fn. 1. Adams (1993) provides a nice summary of the argument.

Now it is not entirely clear (to me anyway) what makes a difference between two relations a conceptual or non-conceptual difference. Adams, also puzzled, suggests that it must be that a difference between relations is conceptual when the relations themselves are conceptual, and otherwise not (see his response to Fodor below). It is also not always clear how the examples Fodor provides to support (2) are supposed to go. For when we spell them out, they don't seem to show what Fodor would like them to. (His example of the property of being a planet, which is supposed to be both relational and causal, will be discussed at length in Ch. 3.)

For example, Fodor argues that the difference between Fodor's having siblings and his Twin's not having siblings is responsible for the difference between his having sons who are nephews and his Twin's having sons who are not nephews. But since the difference in the effects is conceptually connected to the difference in the causes, the property of having siblings is not *causally* responsible for having sons who are nephews. Similarly, in the long quotation above he claims that the property of having had contact with water is responsible for the fact that you have water-thoughts and your Twin does not. But since it is conceptually necessary that if you are connected to water (rather than twater) in the right way then you have water thoughts (rather than twater thoughts), then presumably it follows that having had contact with water is not causally responsible for your having water-thoughts.

But this seems odd. On the one hand it seems to rule out a naturalized or causal theory of content. On the other hand, the individuation conditions of a given type of content – what it is to be a water-thought – are not causal powers; the content itself is a

causal power (or not, as the case may be). As Burge might say, the individuation conditions can be taken for granted when a property is cited in explanation.

So in order to proceed I will simply grant that Fodor's argument makes sense. Thus, to use his other example, while water thoughts cause water behavior ("drilling for water and the like") and twater thoughts cause twater behavior ("drilling for twater and the like"), the difference between having water thoughts and having twater thoughts is not a difference in causal power. This is because it is conceptually necessary that people who have water thoughts (rather than twater thoughts) produce water behavior (rather than twater behavior): "Being water behavior (rather than twater behavior) *just is* being behavior that is caused by water thoughts (rather than twater thoughts)." So being a water thinker is the *same* causal power as being a twater thinker, just instantiated in persons with different causal histories. So a wide content taxonomy is not a taxonomy for causal-explanatory purposes.

Peacocke's response. Peacocke argues that the Modal argument does not threaten wide contents because Fodor misidentifies their explananda. According to Peacocke, wide intentional states explain the relational properties of actions: a movement towards the person in the garden, a movement towards the window, and so on. I pointed to the person in the garden (and not, say, northwards) because I wanted to pick out that person.

He claims that this view of the explananda of externalist states escapes Fodor's argument:

[W]e must distinguish the relational properties of behaviour which I have been emphasizing from the intentions which produced the behaviour. Behaviour which involves a certain relation to water – a relation to something in the external world – does not fall within the scope of Fodor's argument. This is because it is not conceptually necessary that people who have water thoughts produce behaviour which involves relations to water. They will not do so, for instance, if transported (unbeknownst to them) from their normal environment; nor will they do so if

tricked in their normal environment with substances which are superficially water-like.¹⁵²

Peacocke is surely correct in this. My having water-thoughts by no means entails that my subsequent behavior can be truly described as “water-behavior” in the sense of an action that involves water. It can get that description only by virtue of what is the case in the world, and the world is such that my intentions are only contingently satisfied.

But I think Peacocke misidentifies the problem. Even if Peacocke is correct that we explain some relational properties of our actions by reference to wide content, if Fodor’s conceptual connection argument works then we will not have explained how relationally described actions that differ in the way that Twin behaviors do (a movement towards the garden vs. a movement towards the “Twarden”, which would differ in the usual thought-experimental way) should count as differences that matter for causal-explanatory purposes, hence why their causes should count as distinct causal powers. Movements that differ in this one respect do so in virtue of intentions that also differ in this one respect. Fodor challenges us to explain why we should think that these respects matter for causal-explanatory purposes, since he claims that they don’t because of the conceptual connections between them. Peacocke’s response doesn’t answer this.

Adams’s response. Adams (1993) argues that the Modal argument fails to show that there is a threatening type of conceptual connection between wide contents and behavior on either view of intentional action, and that anyway we have no reason to accept Condition C (in which a difference between causes is a causal-power difference “only when it is not a conceptual truth that causes which differ in that one has CP1 where the other has CP2 have effects that differ in that one has EP1 where the other has EP2.”)

¹⁵² Peacocke (1993: 204).

First, on the component view of intentional action, water-behavior is a process which includes a water-intention and a bodily movement as components. On this view the water-behavior – the whole – has this description because of its component water-intention; there can be no water-behavior without this water-intention as a part. Twater-behaviors are similarly analyzed but differ from water-behaviors just in virtue of the different component intentions.

On this view of action, there is a conceptual connection between intending to drink water and intentional water-drinking behavior, for the whole action gets its description in virtue of the description that applies to the part. But so what? The conceptual connection between water-intentions and water-behaviors applies to a part-whole relation, not the relation between water-intentions and the bodily movements they cause. Water-thoughts don't cause intentional water-drinking behavior because the water-thoughts are a *part* of the water-drinking behavior. But the relation between the intention to drink water and the resulting movement is contingent. The same goes for twater-thoughts and bodily movements. Yet there is no reason to think that the difference between two contingent connections is conceptual.

For the record, I think there is another non-threatening conceptual connection involved in the etiology of behavior. Beliefs and other presumably causally efficacious mental states share some of the concepts employed in descriptions of behavior and their component intentions. But this is just to be expected if Mentalese is compositional and if intentional behavior is largely rational. Elsewhere, Fodor suggests typical intentional generalizations of this sort:

If you want to *make an omelette*, and you believe that you can't *make an omelette* unless you *break some eggs*, then, *ceteris paribus*, you will perform an action that

is intended *to be an egg breaking*. (Whether the action actually succeeds in being an egg breaking depends, of course, on whether the world cooperates and the egg breaks.)¹⁵³

Now there is a conceptual connection between desiring an egg-based food item (e.g. an omelette) and intending to break an egg (as well as breaking one). For both mental events share the same concepts. The same would go for Twin-egg-thoughts and Twin-egg-breaking intentions. In each case there is a conceptual connection between the desire and the intention that is due to the sharing of concepts, but even if the difference between them is thereby conceptual this should not undermine a causal relation between them.

Second, on the “product” view of action, which (Adams claims, and I agree) Fodor appears to presuppose, an action is a bodily movement that has an intentional description in virtue of the fact that it is caused by a water-intention. A bodily motion would not be an action (or have an intentional description) if it did not have that etiology. Now it is true that *given* that I intentionally drink water, I must have had the intention to drink water. For one cannot intentionally A (perform a bodily movement that counts as an intentional A-ing) without intending to A. So intentional water-behavior cannot fail to have a water-intention as a cause; it gets that description because of the intention that causes it. But water-intentions could fail to cause water-behavior; they could fail to cause anything at all. So that relation is contingent. The same goes for twater-intentions and twater-behaviors. Again, since each individual connection (between water-thoughts and behaviors, and twater-thoughts and behaviors) is contingent, there is no reason to think that the difference between two contingent connections is conceptual.

Thus, on both views of intentional action, Fodor’s argument “evaporates”. Adams also thinks that we have no reason to accept Condition C to begin with. He shows this

¹⁵³ Fodor (1994: 4).

with the following case. Given that I intentionally move my left arm, I must have had the intention to move it; ditto for my right arm. It is also conceptually necessary that people who have left-arm thoughts (rather than right-arm thoughts) produce left-arm (rather than right-arm) behavior. If Fodor's Condition C were accepted, we would have to conclude that due to these conceptual connections my intention to move my right arm is the same causal power as my intention to move my left arm. This result is "not good!"

My Response. I think Adams provides an adequate response to Fodor's argument. The only thing I have to add is showing the role that is being played by a background assumption that physical identity determines causal-power identity. This issue will arise again in Ch. 3 in my discussion of Fodor's Non-Locality argument.

We may imagine the following response to Adams: There is an essential difference between water- and twater-behaviors (and their causes) and right- and left- arm movements (and their causes) that makes the cases disanalogous. The arm movements are physically distinct, and an intention to move the right arm rather than the left has other distinct properties, like being connected to right-arm muscles and to distinct proprioceptive informational states. But in Twin cases the intentional descriptions of the movements Twins make are their *only* difference. Otherwise they are physically identical.

A relevant analogy would be the following. The difference between a sunburn and a Twinburn (a physically identical burn caused by a Twin sun) is in their etiologies. The sun's power to cause a sunburn is nothing over and above its power to cause a burn. The Twinsun's power to cause a Twinburn is nothing over and above its power to cause a burn. Both burns are physically identical. It would be silly to classify the difference between these effects as relevant for causal-explanatory purposes. So presumably it

would also be silly to classify the sun's and Twinsun's causal powers as distinct.

Similarly, a water-intention's power to cause water-drinking is nothing over and above its power to cause bodily movement B. A twater-intention's power to cause twater-drinking is nothing over and above its power to cause bodily movement B. These bodily movements are identical. It would be equally silly to classify them as distinct effects for causal explanatory purposes. But then their causes, which are also physically identical, should also be classified as the same causal powers.

In short, I think Adams does provide an effective response to Fodor's Modal argument, but also that Fodor has another one up his sleeve which does not rely on claims about differences between relations being conceptually related. I'll return to this argument in Ch. 3.

Summary. Fodor argues that the difference between water-thoughts and twater-thoughts is not a difference in causal power. The difference between them is not a difference in causal powers because the difference in their effects is conceptually connected to the difference in the causes. They are the same causal powers in virtue of a narrow intentional state that they both share.

Adams responds that Fodor's argument "evaporates" on both the component and "product" views of action. In the first case, there is a conceptual connection between an intention and an action, but it applies to a part-whole relation, not a causal relation. On the product view, it is necessary that if there is an intentional A-ing, then it must have been caused by an intention to A. But the intention to A is only contingently connected to the A-ing, since the intention might not have caused anything at all. Thus, the only

connections in each case that seem to matter are contingent, and there is no reason to think that a difference between two contingent connections will be conceptual.

Adams also argued that we have no reason to accept Fodor's Condition C, the no-conceptual-connection requirement. For that condition would require us to accept that the right- and left-arm moving intentions are the same causal powers, which seems absurd. Fodor might respond that the right/left arm case is disanalogous to the water/twater case, since in the latter case are the movements (and intentional states) physically identical. I suggested that the idea that physical identity determines causal-power identity may be lurking up Fodor's sleeve. I discuss this issue further in the next chapter.

2.5 CHAPTER SUMMARY

In this chapter I have presented a series of arguments against the causal efficacy of externalistic content:

(1) Fodor's Argument from the Formality Condition claimed that a scientific psychology that wants to use externalistic states in its explanations would have to wait until all the other sciences have finished their classifications before it could formulate its laws. This argument fails because it confuses an account of the conditions for having a special-science property with an account of that property's explanatory role. We can use contents in causal explanation without having to do a theory of content first.

(2) The Arguments from Narrow Explananda claim that psychology should not use externalistic states in its causal explanations because it ought to individuate its explananda narrowly. Although these arguments are prescriptive – in theory, they leave

open whether wide contents are causes as a matter of fact – they raise the possibility that, despite appearances, folk psychology is not in the business of genuine causal explanation.

Stich's Replacement argument centered on the claim that the sort of systematicity relevant for causal explanation requires autonomously described behavior. I argued that he does not show that autonomous behavioral descriptions are systematic in virtue of their autonomy. So we are given no reason to think that being autonomous is necessary for being systematic. (Presumably it is not sufficient either: we can imagine autonomous descriptions that are unsystematic.) Stich's argument is thus essentially incomplete.

Kim's Causal Core argument claims that it is in virtue of their internal causal cores that externalistic states play the causal role in psychological explanation that they do. I argued that this argument rests on an unstated criterion for causal efficacy that is assumed to leave an internal causal core. But other criteria would leave a causal core that is partly external. Thus, Kim needs to argue for his criterion before he can use it in his argument. But once he does that, the Causal Core argument is superfluous. In short, Kim's argument is also essentially incomplete.

I suggested that the sorts of causal intuitions that Kim and Stich appear to be relying on implicitly are spelled out in the Non-Locality arguments presented in the next chapter (Fodor's Non-Locality argument in particular).

(3) In the Cross-Context argument, Fodor argued that sameness of causal powers must be individuated across contexts, not within contexts. Burge argued that the value of the test depends entirely on which contexts are considered relevant. We can only infer to sameness of causes from sameness of effects in contexts where a difference of causes

could make a difference if there is one. As Burge points out, ruling out contexts where causal-historical properties might make a difference in effects is question-begging.

(4) In the Modal Argument, Fodor argues that the differences in externally individuated causes and the differences in the descriptions of their effects are conceptually connected. Since causation is a matter of contingent relations, cause properties that differ in this way do not differ as causal powers – they are the same causal powers. Adams argued that on either view of intentional action, the relations between water-intentions and water-movements, and twater-intentions and twater-movements, is contingent. So there is no reason to think that the differences between these relations are conceptually connected. I suggested that Fodor may have another argument up his sleeve, based on the intuition that physical sameness and difference determine sameness and difference of causal powers. This argument will be addressed in Ch. 3.

One interesting result of this survey of arguments is this. We find Kim and Fodor arguing that some *intentional* properties can play a causal role (albeit narrow ones), and Fodor arguing that some *relational* properties can play a causal role (being a planet), yet for some reason the combination of *intentional* + *relational* is utterly fatal to causal efficacy: causally efficacious properties can be intentional or relational *but not both*. Given the standard truth table for “or”, this is a logically invalid conclusion. At least, it cannot be determined by *a priori* argument. So there is reason to think that there is some deep confusion going on in the debate so far. However, the most significant doubts about the causal efficacy of content remain to be addressed.

CHAPTER THREE: ARGUMENTS FROM NON-LOCALITY

3.1. INTRODUCTION

3.1.1 Review

In Chapter 2 I presented and responded to a series of well-known arguments against the causal efficacy of wide content. These arguments have helped to convince many that no robust causal role for wide content is possible.¹⁵⁴ I'll summarize those results before discussing the next (and final) group of arguments, the Arguments from Non-Locality, in this chapter. I consider the Non-Locality arguments fundamental in that they focus directly on the "wide" aspect of wide content. They claim directly that such contents are not causally efficacious for reasons stemming from the fact that they are extrinsic. This claim is often expressed by saying that wide contents are "non-local" (hence my label), although there are different ways of specifying what "non-locality" amounts to. I focus on the arguments provided by Fodor and McGinn.

This chapter also completes the first task of my dissertation, which was to determine whether the various arguments presented in the literature establish that there are genuine problems for intentional causation arising from their extrinsic nature. Those arguments will all have been surveyed by the end of this chapter. If successful, my arguments have shown that we have no a priori reason to bar wide contents from being robust mental causal powers. In Chapter 4 I turn to the second task of this dissertation, which is to provide a scientifically-grounded positive account of how wide contents could be robust causal powers.

¹⁵⁴For example, Fodor (1994) is motivated by the claim (on the book's back cover) that "[a]mong philosophers, a general consensus exists that a referential semantics for mental representation cannot support a robust account of intentional explanation."

I divided the arguments presented in Chapter 2 into two main groups: Arguments from Narrow Explananda and the Cross-Context and Modal Arguments. In the former group I included Stich's Replacement Argument, which argued from the goal of systematic causal explanation of Twin behaviors to the need for internally individuated causes. I also included Kim's Causal Core Argument, which argued from the existence of a narrowly individuated causal core in many widely individuated states to the causal irrelevance of the relational features that are left out of the core. Each of these arguments represented a class of similar objections.

I responded that both arguments ultimately rest on undefended hidden premises. The Replacement Argument relies on the claim that causal explanations citing narrow properties are systematic because the properties are narrow. But Stich does not provide an argument linking the causal-explanatory systematicity of narrow properties (or narrowly individuated states) to their narrowness. The Causal Core Argument relied on the claim that causal cores will be narrow. But Kim does not provide a non-question-begging principle for "carving out" causal cores that does not rule out extrinsic properties as causes by definition. I also suggested that the gap in each argument could be filled by an argument linking the alleged causal inefficacy of wide contents directly to their extrinsic character – in short, a Non-Locality argument. For if wide contents are causally inefficacious *because* they are extrinsic, this fact would entail that Kim's causal core will be narrow and would strongly support Stich's claim that the narrowness of syntactic properties is the source of their causal-explanatory systematicity.

Next I discussed Fodor's Cross-Context and Modal Arguments. Both essentially involve our acceptance of certain counterfactual causal judgments that support the claim

that differences between wide contents are not differences in causal powers. (If they are the same causal powers, then the extrinsic features by which they differ play no part in assuring their status as causal powers.) The Cross-Context Argument claims that when we compare causal powers across contexts, wide causal powers will not differ across different contexts. The Modal Argument claims that the differences between the effects of different wide contents are conceptually connected, and thus not causally connected, to the differences in their widely individuated causes.

In response, I argued that the Cross-Context Test is only valid as a diagnostic of differences in causal powers in contexts in which a difference between causal powers that are due to relational differences could appear if there is one; otherwise the “test”, and the argument, is inconclusive. I argued that the Modal Argument involves two distinct sorts of conceptual connection – between contents that contain the same concepts, and between the alleged causes and effects. The latter connection, if true, is a serious problem, but it is not new to Fodor, nor is it limited to wide contents. But I concluded that this result, along with Kim’s argument, motivates an account of how wide contents can be causally efficacious in singular causal interactions, in particular between token mental states and token basic actions. That is, it motivates an account of robust semantic causal efficacy.

3.1.2. General Remarks on Arguments from Non-Locality.

What I am calling Arguments from Non-Locality all start from claims about the “locality” of causation, which refers ultimately to the spatiotemporal features associated with mechanistic causation. I will explain below how term “locality” is used ambiguously, but intuitions regarding mechanistic causal interactions are the starting

point and the driving force of the objections. Kim alludes to such features when he cites the “considerations of causal continuity and contiguity” that make it plausible to think that the causal core of any wide state will be narrow (see 2.3.2). Similarly, Stich (1983: 64-65) states in his Replacement Argument that causal history and sociolinguistic environment “are irrelevant to a psychological theory except insofar as these differences make a difference to the organism’s current, internal, physical state.” These remarks indicate that Non-Locality arguments are implicitly bridging the gaps I identified in the Replacement and Causal Core arguments. They therefore play a fundamental role in the overall debate, one that justifies McGinn’s (1989) assessment of the Non-Locality arguments as “the most serious and challenging argument[s] against content-based psychology”.

Stich and Kim do not develop Non-Locality arguments, but their remarks reflect the basic strategy: to conclude that extrinsic properties do not play robust causal roles from premises that start from the claim that causation is “local” in the sense that there is no action at a distance in time or space.¹⁵⁵ The intuitive idea of “local” causation is that there can be no gap in space or time between a cause and its direct effects.¹⁵⁶ So the claim is that externalistic properties violate this physical desideratum. Non-Locality arguments fill out the steps in this inference by linking the extrinsic nature of wide contents to a kind of “non-locality” and from there to their causal inefficacy.

¹⁵⁵ For example, Block (1986: 625) argues against wide contents as causes because differences in thoughts due to relational differences “do not cause any differences in proteins, synaptic connections, or any other physiochemical feature of the body, as do differences in genes and surface stimulations. This would amount to a kind of action at a distance, and this would clearly go counter to the methodology of attribution of variance.” (I argued against the *a priori* truth of the first sentence of this quote in 1.5.2.) Pettit (1986: 35-36) makes a similar point about action at a distance in the explanation of intentional action.

¹⁵⁶ Lange (2002) has an excellent discussion of the philosophical problems involved in “action by contact”.

In what follows, I first discuss McGinn's "referential" Non-Locality problem (related "referential" arguments by Pylyshyn and Fodor are also presented). My claim is that McGinn's argument is based on a misconception about wide contents that is rooted in the relationality conception of "extrinsic". He claims that instances of wide content properties must have microphysical bases that are partly spatiotemporally remote, thus violating the requirements of local causal interaction. I reply that externalism does not entail that instances of extrinsic properties must have spatially remote microphysical bases. The externalist can hold that such properties are wholly constituted in physical (at some level) properties located in the head as long as their constitution is not sufficient to determine their identity. The essential difference between extrinsic and intrinsic properties lies not in the location of their microphysical bases, but in what a token's having them implies about that token's world. If P is an extrinsic property of x, then x must be in a world which has a certain character – for example, it contains H₂O. If P is an intrinsic property of x, it entails nothing about what x's world must be like; x's world need not have any particular character at all.

Next, I discuss Fodor's version of the problem, which I consider the strongest. (A related argument by Dretske is also presented.) This Non-Locality argument takes its cue from the fact that different mental properties can have identical "local" components; for example, Twins have the same "local microstructure". Fodor claims it is scientifically absurd to distinguish the causal powers of such properties. I reply that the validity of his argument depends on whether "local microstructure" is extrinsic or intrinsic. I do not claim that the extrinsic hypothesis is true, but I provide reasons to think it might be. The important point is that there is no *a priori* reason for thinking that wide properties cannot

be robust causal powers. Our intuition that the contents of our thoughts do make a causal difference, and that ordinary psychological explanation is genuinely causal explanation, is *a priori* justified and is as likely as not to be supported empirically.

I close the chapter with a final argument from Fodor and an extended discussion of his defense of the causal efficacy of the property of being a planet. I show how Fodor does not succeed in distinguishing those extrinsic properties that he claims are causal (being a planet) from those that by his lights are not (being a water-belief). That is, if his arguments against beliefs were successful, they would also succeed against planets; while if his defenses of planets were successful, they would also succeed for beliefs.

3.3. ARGUMENTS FROM “REFERENTIAL” NON-LOCALITY

The first sort of Non-Locality argument I’ll examine presents the problem in terms of the representational or “referential” nature of content. Arguments of this sort stem from the traditional view that intentionality in general involves a relation or attitude towards a mind-independent entity, a proposition. Relations to these entities will be external, and possibly (depending on one’s view of propositions) non-natural. However, since referential relations to propositions are true of *all* contentful states (assuming there are propositions), the “referential” versions do not target externalistically individuated contents in the sense defined in 1.2.2 in particular. Nor do they presuppose semantic naturalism. However, a special case of the “referential” version arises if we assume semantic naturalism, as this quote from Field helps to show:

I first formulate Brentano’s problem as the problem of making naturalistic sense of belief, desire, and so forth, given that they are relations between organisms and intentional entities like propositions. ... [W]e might regard propositions as useful fictions (or maybe even not-so-useful fictions). In that case there is no pressure to

solve Brentano's problem *as I stated it*, i.e. in terms of propositions. But this wouldn't dissolve questions about how mental states can stand in referential relations to the world. (2001: 69)¹⁵⁷

Since I have assumed semantic naturalism (see 1.2.1), my remarks regarding McGinn's argument will apply to this special case, which is the dominant contemporary view. I also previously limited the scope of my discussion to mental states that are uncontroversially about existing objects and features in the world (see 1.2.1). The issues of reference involving non-existents, abstract objects and other non-concrete items will be addressed in my discussion below of Pylyshyn and Fodor, who specifically question the causal efficacy of mental states that represent non-concrete items.

3.3.1. McGinn's Argument from "Referential" Non-Locality

Argument. McGinn states a Non-Locality argument in terms of the referential relations between words and the world that underlie a correspondence theory of truth:

According to externalism, contentful states are identified by reference to entities that lie outside the subject's body (including, of course, his head): these extrinsic entities are the constituents of such states. Now this implies that contentful states are (as we might say) extrinsically relational: they essentially consist in relations to extrinsic nonmental items (objects, properties, etc.). They thus involve certain 'correspondence relations' between the subject (or his intrinsic states) and entities that exist beyond the confines of the subject. For example, for John to have the concept *square* (and so to be capable of believing that the box is square) is to stand in a certain relation to the *property* of being square. So when we ascribe that concept to John we report him as standing in this extrinsic correspondence relation to that objective property. And now the causal objection to citing such contents in psychological explanations is just this: if such explanations purport to be causal, then these correspondence relations cannot themselves be implicated in the causal transaction being reported. This is because what happens at the causal nexus is local, proximate and intrinsic: the features of the cause that lead to the effect must be right there where the causal interaction takes place. Causation is the same with brains and minds as it is with billiard balls. Their effects depend on local properties of these entities. The causal powers of a state or property must be

¹⁵⁷ The reference to his "first" formulation is to Field (1978), which laid the groundwork for transposing the traditional view of mental states as attitudes to propositions to the current naturalistic orthodoxy (see 1.2.1).

intrinsically grounded; they cannot depend essentially upon relations to what lies quite elsewhere. (1989: 133)

He adds that the “causal lack” (i.e. inefficacy) of wide contents can be traced to the “relationality of the properties in question.” Presumably the representational properties in a deflationary theory of content that involve only “indication” relations, not full-blooded “reference” relations, would be equally vulnerable to this argument.¹⁵⁸

Without yet analyzing the argument, I should note that McGinn does not even attempt to defend a robust causal role for content in the face of it. Instead, he focuses on justifying our use of content-involving explanations *given* that contents are not robustly causal. This justification involves adopting a teleological or teleo-informational theory of content, in which content properties classify mental states by adaptive function defined relative to a given environment (see 1.7.3). The justification is that our content-based mode of explanation explains the adaptive functions that the causal mechanisms subserve. Mental states are identified in terms of extrinsic worldly descriptions because these descriptions relate the subject functionally to her environment. This is consistent with the externalist’s claim that content can differ when intrinsic states do not, since looking at what’s going on inside won’t reveal the functions that are being subserved by what goes on inside.

This response does provide for a good explanatory role for content – as does Dretske’s account of structuring causes (discussed below). But it is inadequate from the point of view of anyone (such as myself) who wants to defend a robust causal role for wide content and not just any explanatory role. Contents should be causal-explanatory

¹⁵⁸ For example, see Field (1994) for a deflationary theory of meaning and content. As noted in 1.5, this generalizing of McGinn’s argument is easily made on the view of extrinsic property I prefer, but is not possible on the relationality view, where a difference in relation entails a different relational property. But surely McGinn’s argument should apply to both cases without having to specify their differences.

because they are causally efficacious, just as other causal-explanatory properties are. However, unlike McGinn, I can find no argument in this passage that is worth worrying about. The basic problem is that McGinn appears to think that from the fact that relations to external relata are essential to (“constitutive of”) extrinsic types, it follows that external relata are literally constituents of the tokens of extrinsic types. Those who think McGinn *can't* be making *this* mistake may substitute the name “McGoo” for any token of “McGinn” in what follows; “McGoo” will refer to a fictional philosopher who has possessed McGinn and used him as an unwilling mouthpiece.

My Response. To begin with, the term “local” is being used in the above quotation in several different, logically independent senses. One implicit application of the term is to causal interactions: causal interactions are local iff they occur by direct contact between individuals, such as two billiard balls. This is the locality of causation – the physical desideratum – that generates the argument and is the source of its intuitive strength. Externalists and their opponents alike agree that causal interactions are local in our world. For all intents and purposes, the subset of logically or metaphysically possible worlds in which causal interactions are local is the set of physically possible worlds, relative to the physical possibilities established in our world.

But the term “local” is also applied to properties, not causal interactions. In the above quote, McGinn is charitably interpreted as referring to type essences: “locality” is a matter of what is essential to or ‘constitutive of’ a property. So a property is “local” or not in this sense depending on its nature – that is, depending on the conditions that must be met for a particular to count as a token of that type. Properties that involve relations as part of their nature are not “local”; those that do are “local”. That is, McGinn assumes a

relationality conception of “extrinsic”. In this type-essence sense, “local” is simply synonymous with “intrinsic”.

Unfortunately, while this reading may be charitable, it leaves McGinn empty-handed. An argument to the conclusion that extrinsic properties can’t be causal powers because they are not “local” in this sense is not an argument at all. On this reading, the above passage is just a complex, emphatic way of restating the simple conviction that only intrinsic properties can be causal powers. For to say that extrinsic properties can’t be causal powers because they’re not “local” in the type-essence sense is just to say that they can’t be causal powers because they’re not intrinsic.

There *is* an argument here, however, if we expose a third sense of “locality” that is hinted at in the claim that extrinsic entities are the constituents of extrinsic states. This application of the term “local” is derived from a feature of the material or spatiotemporal constituents of paradigmatic concrete particulars. In this sense, a property-instance is “local” if it or its microphysical base lies entirely within the spatial boundary of a concrete individual.¹⁵⁹ Property instances are “local” derivatively, since fundamentally only tokens have spatiotemporal location.¹⁶⁰

If externalism entailed that content properties were “non-local” in the property-instance sense, then it might be argued that such properties can’t be causal powers because their microphysical bases are not wholly there where a token “local” causal interaction must take place. As McGinn puts it above, they would not be “intrinsically

¹⁵⁹ Odd cases of discontinuous boundaries include bikinis and universities, but these subtleties won’t matter here. Also, the idea that all physical properties (let alone special science properties) can be identified with mereological composites of more basic microphysical properties is a basic part of physicalism, discussed in Kim (1993: 101-102). The claim that higher-level properties are constituted of but not identical to these mereological composites is defended in Pereboom (2002), which I discuss at length in Chapter 4.

¹⁶⁰ Trope theorists would say tropes (particular attributes, e.g. this ball’s red color) are “local” in this sense.

grounded.” In short, armed with this third sense of “locality”, the above passage might be expressing the following argument:

- 1) Only locally instantiated properties can play genuine causal roles in local singular causal interactions.
- 2) Externalism entails that wide contents are non-locally instantiated.
- 3) So wide contents cannot play genuine causal roles.

Although the above quotation alone does not contain this argument, I think it can be extracted from McGinn when the citation is combined with his analysis of externalism and what it entails. After defending the claim that McGinn does in fact hold premise 2, I’ll be arguing that premise 2 is false and that the argument is unsound.

McGinn regards externalism as a case of asymmetrical individuation-dependence of Fs on Gs, such as that of sensations on persons.¹⁶¹ This dependence has four aspects – linguistic, epistemological, metaphysical and conceptual, of which only the metaphysical is of concern here. The metaphysical aspect of individuation-dependence holds that “the essence of Fs is (partly) constituted by that of Gs.” He explains:

Thus, taking the example of sensations and persons, we might want to say the following (assuming this is an instance of the general notion [of individuation-dependence]: ... (iii) it is part of the essence of a (token) sensation that it be enjoyed by a particular person; ... Now how does this general schema fit the case of externalism? ... I think it fits rather neatly. Let the Fs be beliefs and the Gs the objects and properties environmentally related to these beliefs, as it might be causally. Then, in holding that these beliefs are individuated by the environment, the externalist is advancing the following [subthesis]. (iii) It is of the essence of a particular belief that it be environmentally related to certain entities; indeed its very constituents are such worldly entities. It is simply not *possible* for the belief to be held by a subject unless his environment contains the appropriate entities, those the belief purports to be about. (1989: 4-5)¹⁶²

¹⁶¹ He notes that the dependence will typically be partial, a necessary but not sufficient condition of identity for Fs in terms of Gs. So while McGinn does assume the relationality conception of “extrinsic”, he does not identify relational properties with relations.

¹⁶² McGinn (1989: 28) also attributes a failure of epistemic access to content on this non-locality: “As I have observed, externalism inserts spatially distant objects into the content of thoughts. Such thoughts will not be transparent, because the identity and existence of their constituents will not be evident to introspection. ... Opacity of content goes with the location of the constituent: if the constituent comes from the external spatial world, then the content of the thought will be opaque (or at least not fully transparent).”

A charitable interpretation of this passage holds that McGinn is talking about type-essences only and what externalism claims is essential to mental state types. The last sentence in particular expresses very nicely the conception of “extrinsic” that I have defended. However, while McGinn otherwise assumes the relationality conception of “extrinsic”, I can agree for the sake of argument that part of the essence of a given belief-state type is that it be “constituted by” a relation to the environment.

But given the explicit reference to tokens, it is difficult to avoid a reading in which McGinn slides from talk of what is “constitutive of” mental type-essences into talk of belief tokens and from there into what constitutes a belief token (or a token content).

The slide also seems to occur in the following claim:

We should reject the idea that it is some unreconstructed brand of materialism that gives off a reluctance to countenance externalism. It is true that identifying the mind with the brain, and mental states with brain states, conflicts with externalism, since the brain really does lie literally within the head. So it is a consequence of externalism that this simple kind of materialism is false. (1989: 15)

This passage can also be read as being about mental state types: externalism does indeed entail that extrinsic mental state types cannot be identical to brain state types. But this reading must overlook the reason McGinn gives for the failure of identity: “since the brain really does lie literally within the head”. Token brains (hence token brain states) do lie literally in token heads, but *that* isn’t why brain state types aren’t identical to extrinsic mental state types. By analogy, the type “cat” cannot be identical to the type “bikini”, but the fact that token cats, but not token bikinis, drown when underwater is not why type-type identity fails. Similarly, the type “water-belief” cannot be identical to the type “neural state N”, but the fact that tokens of neural states actually lie literally within heads

has nothing to do with the failure of type-type identity. Lying literally in the head may not even be an essential property of neural state tokens, nevermind neural state types.

An even more charitable reading simply omits the bit about lying in the head. But there are plenty of other passages where it is very difficult to dismiss the suspicion that McGinn has confused what is constitutive of a type with what constitutes its tokens. He claims that “full-blooded externalism can and should be formulated in terms of the notion of external constituents”; that externalism “inserts spatially distant objects into the content of thoughts”; that “externalism entails that minds are not substances” because minds “must traverse space – reach out across it – in order to have spatially located constituents”; that externalism “involves ontological dependence and breached boundaries (or no viable notion of boundary at all) – exactly contrary to substantialism”; in a word, that externalism violates our notion of substance:

Substances are spatial beings whose configuration is independent of what lies beyond them; they do not incorporate constituents from elsewhere. That is essentially what their self-subsistence consists in. ... Told of some putative entity that had spatially remote constituents upon which it was ontologically dependent, we would conclude that the entity in question could not be a substance – that it could not be of the same metaphysical category as rocks and cats and kidneys. But this is precisely what externalism affirms about the mind: it is constituted by its relations to distant objects. The characteristic properties of mind – its contentful states – are therefore not like intrinsic primary qualities of material substances, but are rather extrinsic relations that may take as their relata items from elsewhere in space. Neither does the mind seem to have clear (or clearly intelligible) boundaries, according to externalism. (1989: 21).¹⁶³

On McGinn’s behalf, I should note that there are philosophers who explicitly defend this type of externalism, which (in 1.2.3) I called “vehicle externalism” and which also goes by the name of the Extended Mind thesis.¹⁶⁴ In Clark and Chalmers (1998), it is the claim

¹⁶³ These quotations prior to the long citation are at (1989: 43, 21, 16-17).

¹⁶⁴ Egan (unpublished MS) distinguishes between “taxonomic externalism” – which involves individuation – and “constitutive externalism” – which involves instantiation – and argues that while the former may or

that there are “coupled” or extended cognitive processes that literally include objects outside the head – such as pen and paper, slide rules, Palm Pilots, and so forth. So in ascribing this view to McGinn, I am not saddling him with a view that has no other defenders.

But even explicit proponents of the Extended Mind thesis do not claim that content externalism *entails* the Extended Mind thesis; they treat them as separable theses that require distinct arguments.¹⁶⁵ McGinn does not provide any argument for it, and *pace* his remarks, it does not simply follow from externalism. As far as I can tell, these passages generally involve a failure to distinguish between the following:

(A) A token water-belief must exist in a world where water samples exist.

(B) A token water-belief must include water samples as constituents.

(B) entails (A), but not vice versa. But externalistic individuation only entails (A); it does not entail (B). The fact that water must exist in the token’s environment is not the same as the fact that water must be a constituent of the token – even if it is true (on a different understanding of ‘constituent’) that it is constitutive of a belief-type that it involve relations to water. Moreover, it need not be essential to a token water-belief that it be a water-belief, even if it is essential to its being a water belief that there be water instantiated in the token’s environment. (A) and (B) state (differing) claims about what is essential for x’s being a water-belief, but nothing about whether being a water-belief is an

may not be true, the latter is not. (I agree.) She lists among the proponents of the *Extended Mind* thesis Andy Clark, David Chalmers, Daniel Dennett, John Haugeland, Susan Hurley and Robert Wilson. For the record, Egan is a “taxonomic” internalist.

¹⁶⁵ For Clark and Chalmers, the fact that the environment constrains the evolution and development of cognition supports the idea that cognition is extended. (Note how this claim is reminiscent of the physical conception of property as a constraint that I suggested in 1.5.1). Egan (unpub.) responds that the adaptive mechanisms that do the exploiting are located entirely in the head. I agree with Egan, but if my account in Chapter 4 is plausible, it doesn’t follow that these mechanisms are intrinsic.

essential property of x. It might be, but externalism doesn't entail anything about whether extrinsic properties are essential to the particulars that have them.

In short, externalism *does* have consequences regarding constituency, but not regarding the material constituents of mental tokens – only regarding the constituents of the world in which a mental token exists. It is compatible with externalism that instances of extrinsic properties are fully constituted in microphysical properties in brains – that their microphysical bases are as “local” as the microphysical bases of intrinsic properties. What externalism forbids is that they be identified with these microphysical bases. It is open for the externalist to affirm a constitution thesis between mental tokens and physical tokens, or between mental property instances and microphysical property instances.

McGinn claims that this constitution move is not possible for the “physical basis” of mental states. Analyzing the consequences of externalism for physicalism, he writes:

The contours of the physical bases of mental states must be moulded according to the shape of the mental states so based. My question, then, is how the physical basis must be sculpted in order to provide for externalism. What seems immediately clear is that the physical basis cannot be located wholly within the head. We cannot site the mind at the place of the brain, and we cannot make intrinsic brain states the (full) basis of representational mental states. (1989: 103)

The failure of identity, he claims, is due to a failure of Leibniz's Law: “x's mind is essentially not a substance, whereas x's brain essentially is a substance.” Whether identity fails for this reason or not, we can grant for the sake of argument that it does fail.

But he then adds:

Nor will retreating to a constitution thesis help: for if brain states wholly constitute mental states, then it looks as if mental states will have to exist in any world in which their constitutive brain states exist – but this will be false of the worlds in which the objects of the mental states do not exist. What has been omitted are the extrinsic relations definitive of the mental states. (1989: 104)

However, this argument is obviously question-begging. Anyone who holds that mental states are wholly constituted of brain states but not identical to them will deny that in any world in which the constitutive brain states exist the mental states do – just as anyone who holds that a statue is wholly constituted in a piece of marble yet is not identical to that marble piece will deny that if the marble piece exists in a world the statue does too. So token beliefs can be literally in the head even if their identity is not determined by what is in the head. Instances of extrinsic properties are individuation-dependent on the environment, but they need not be constitutionally dependent on the environment.

McGinn's mistake – failing to distinguish between "local" property instances and "local" property essences – can also be illustrated by the case of a pencil. The conditions that must be met for a token to count as a pencil include being made or regularly used in accordance with human intentions; so being a pencil is not a "local" property in the type-essence sense. But each instance of the property of being a pencil can be "local" in the property-instance sense if the property instance is constituted in but not identical to its microphysical base. The fact that having that microphysical base is necessary but not sufficient for being a pencil does not imply that what's missing are *more microphysical parts of the pencil*. A property-instance of being a pencil is an extrinsic property-instance because a token's having that property implies that it exists in a world where minded beings exist. A token's having that property does *not* imply that the token must have spatiotemporal constituents outside that spatial boundary. So pencils are instantiated "locally" in the property-instance sense of "local" and individuated "non-locally" in the property-essence sense of "local".

I conclude that on the only reading of McGinn in which he presents an *argument* – as opposed to a mere restatement of the conviction that only intrinsic properties can be causal powers – the argument rests on a mistake. Pace premise 2 above, externalism doesn't entail anything about the location of the material constituents of token beliefs or the microphysical properties that may constitute the supervenience base of a content property-instance. It does entail something about the constituents of the world in which the token belief is located, which is quite another matter.

In fact, it should be obvious that externalism *cannot* entail that the supervenience bases of mental-property instances contain external relata as constituents. This is because the extrinsicness of the relations and the constituency of the external relata are logically incompatible. If the external relata are constituents, then the relation to them cannot be an extrinsic one; the relata are not external at all. But if the relation is an external one, then the external relata cannot be constituents. Relationality in property essences and constituency in property instances don't mix. That is why I cannot make sense of McGinn's (1989: 39) claim that "since beliefs are relations to these possible states of affairs [which contain worldly objects], they contain whatever the states of affairs they represent contains." If beliefs contain them, the relations can't be external; and if the relations are external, they cannot literally contain them. McGinn wants both relationality and constituency, but he can't have both.

Note that the extrinsic definition I defended in 1.5.1 – in which an extrinsic property of an object is one which entails the existence of an object distinct from its possessor – is highly unlikely to provoke these conceptual errors. From the fact that the instantiation of an extrinsic property by an individual entails that another distinct

individual exists, it clearly does not follow that the extrinsic property is instantiated by both individuals. To say that a property is extrinsic is to conditionalize the sort of world in which its possessor exists; if x has P , then x 's world must be thus. This gives us no reason at all to think that the external object's constitution is part of the microphysical basis of the extrinsic property's instances.

If there is any argument in McGinn beyond the one just analyzed, it would seem to be indicated by this comment:

Surely, if the external relata of these relations were not to exist, this would not necessarily alter the causal power of the states of the subject to bring about the same bodily movements. At best, therefore, the extrinsic relations definitive of content are causally redundant, thus inviting an application of Occam's razor. (1989: 134).

However, the intuition expressed very briefly here – that a subject's mental states that count as different beliefs should count as the same causal powers – is made much more explicit in Fodor's Non-Locality argument, discussed below. But this argument is compatible with the microphysical bases of content property instances being wholly located in the head.

3.3.2. Specific "Referential" Arguments

Other "referential" arguments focus on specific cases of referential relations rather than on the reference relation in general or on relational properties in general. The basic idea is that we can have thoughts about non-existents (Hamlet), abstract objects (numbers) and future states of affairs (the sun's going supernova). The fact that none of these relata are the sorts of things we can be causally related to is supposed to support the claim that contents that are about them must be causally inert.

Pylyshyn's (1984) version of the problem begins with a story of a pedestrian who rushes to call 9-1-1 because, after observing an accident, he believes that there is a need for assistance and that calling 9-1-1 is a good way to get help. The problem with citing the contents of his states as causes of the pedestrian's behavior, Pylyshyn claims, is that his behavior "cannot be explained in terms of a chain of causal laws connecting such extrinsic entities as those mentioned in the account to the current state of the system." The extrinsic entities he mentions include the contents of goals, which are possible states of affairs; the contents of false beliefs; and contents about numbers and nonexistents such as Santa Claus, a hoped-for raise, and so forth. Clearly there are no causal relations between a thinker and these entities. In his terms (also partly cited above):

When we say that we do something because we desire goal G or because we believe P, we are attributing causality to a representation, though it is not the content that does the causing. As we have noted, what is being represented need not exist, and even when it does, its only relevant aspect is the way it is conceptualized. What actually does the causing are certain physical properties of the representational state – but in a way that reflects the representational state's content. (1984: 39)

Fodor (1987) voices a similar problem:

It looks as though intentional properties essentially involve relations between mental states and *merely possible contingencies*. For example, it's plausible that for a thought to have the content THAT SNOW IS BLACK is for that thought to be related, in a certain way, the possible (but nonactual) state of affairs in which snow is black; ... Now, the following metaphysical principle strikes me as plausible: the causal powers of a thing are not affected by its relation to merely possible entities; only relations to *actual* entities affect causal powers. It is, for example, a determinant of my causal powers that I am standing on the brink of a high cliff. But it is *not* a determinant of my causal powers that I am standing on the brink of a possible-but-nonactual high cliff; I can't throw myself off one of *those*, however hard I try. Well, if this metaphysical principle is right, and if it's right that intentional properties essentially involve relations to nonactual objects, then it would follow that intentional properties are not per se determinants of causal powers, hence that there are no intentional mechanisms. (1987: 140-141)

My Response. The only *extra* challenge being raised by these arguments is how to naturalize non-demonstrative non-perceptual contents, a class that includes but is not limited to contents involving reference to possible states of affairs, numbers and non-existents. The arguments only show how embarrassingly inadequate our current causal-informational theories of content are, and that, as they say, is not my problem. For suppose we had such a naturalized account – one in which causal relations are not the core of the explanation or not part of it at all. Perhaps, in empiricist fashion, we could extend our current naturalistic theories by taking perceptual or demonstrative contents as basic and building concepts like Santa Claus, the number two and the rest from them. Or perhaps an internal conceptual-role story is correct for these contents, and they aren't externalistically individuated. Or perhaps some other natural relation will be found that can account for these contents. At the very least, even if the metaphysical principle Fodor cites here is plausible, we've been given no reason to think that the contents of non-demonstrative, non-perceptual beliefs – which would include the problematic sorts specifically mentioned by Fodor and Pylyshyn – cannot be explained consistently with it.

So there is in principle no extra problem being raised here for the causal efficacy of wide content. It is possible that either these problematic sorts of contents are not externalistic, or the external relations upon which they depend are none other than the causal relations that explain the contents of representations that are about actual entities or states of affairs.

3.3. ARGUMENTS FROM SPATIOTEMPORAL NON-LOCALITY

3.3.1. Fodor's Argument from (Spatial) Non-Locality

Argument. Fodor (1987) presents his Argument from Non-Locality as a trump card after conceding that his Cross-Context Argument can be finessed by individuating behavior such that every difference in wide content has a different effect.¹⁶⁶ He begins with the general methodological principle that scientific taxonomies are taxonomies by causal powers, where identity of causal powers is determined by identity of causal consequences across nomologically possible contexts. This plausible methodological principle (which he calls “methodological individualism”) does not rule out extrinsic properties in general as causal powers, nor is it intended to; Fodor claims, for example, that the properties of being a planet and being a meteor are causal powers (but see 3.4 below).

What bars wide contents from being causal powers is the general metaphysical principle that “causal powers must supervene on local microstructure”; as he puts it, “mind/brain supervenience (/identity) is our only plausible account of how mental states could have the causal powers that they do have.”¹⁶⁷ (Why doesn't that rule out “being a planet” as a causal power? See 3.4.) Intuitively, the principle holds that there can be no causal difference without a local difference. If this principle is right, it would seem that the taxonomy of mental states sanctioned by folk psychology would not be a taxonomy

¹⁶⁶ In his (1987: 38) words, “you can save classification by causal powers, come what may, by fiddling with the criteria for event identity” – that is, taxonomizing by causal powers is a trivial constraint if all you need to do to meet it is to individuate effects as distinct whenever you need to.

¹⁶⁷ See *op. cit.*, 44. In section 3.4 below, I discuss in detail all of the various reasons Fodor gives for why (e.g.) being a planet counts as a causal power while (e.g.) being a belief that water is wet is not. It turns out that none of the reasons he gives provides a way to distinguish between these properties.

by causal powers. For on the assumption that brain states are narrowly individuated, the mental states individuated by folk psychology will not supervene on the brain.¹⁶⁸

Fodor's defense of his metaphysical principle is the crux of his Non-Locality argument, which goes as follows. It is a (contingent) feature of our world that causal mechanisms are local; there is no action at a distance. So given the locality of causal mechanisms, the causal powers of thoughts can only differ with differences in external relations in one of two ways: either there would have to be a causal mechanism that mediates the relations, or the differences in causal powers due to these relations would have to be a matter of fundamental natural law.¹⁶⁹ We know there is no such basic law. But we also know that there is no such causal mechanism. For if there *were* such a mechanism, differences in intentional causal powers would be revealed in physiological differences. After all, Fodor (1987: 41-42) exclaims, "how *could* differences in context affect the causal powers of one's mental states without affecting the states of one's brain?" And Twin-Earth thought experiments show that Twin thoughts can differ with differences in external relations *without* there being any corresponding local differences.

If this is right, we face the following choice. We can postulate "impossible causal mechanisms and/or impossible natural laws" to override (so to speak) the failure of local physiological differences to match wide content differences. Or we can conclude, with

¹⁶⁸The operative sort of supervenience is *strong* supervenience, in which there can be no mental difference without a local (i.e. brain or body) physical difference. *Global* supervenience holds that there can be no mental difference without a physical difference in a given world; the difference need not be local. The truth of global supervenience would leave the non-locality problem untouched. I should note, however, that the term "supervenience" is sometimes used ambiguously in this context. For example, Burge defends the causal efficacy of "supervenient" types (which can be strongly or globally supervenient), while Fodor targets "non-supervenient" types (those which violate strong supervenience). The difference should be clear in the text.

¹⁶⁹Mediation by mechanisms is at the token level of singular causal relations, while mediation by laws involves relations between types that subsume the token-level relations. In case of fundamental laws, Fodor's idea seems to be that causal powers would differ with external relations if wide contents played a role in a *new* fundamental law, since no known physical law violates mechanism (oddities of quantum mechanics aside).

Fodor, that the external relations that distinguish wide contents do not make a difference to intentional causal powers:

You can affect the relational properties of things in all sorts of ways – including by stipulation. But for one thing to affect the causal powers of another, there must be a mediating law or mechanism. It's a mystery what this could be in the Twin (or Oscar) cases; not surprisingly, since it's surely plausible that the only mechanisms that can mediate environmental effects on the causal powers of mental states are neurological. The way to avoid making this mystery is to count the mental states – and, *mutatis mutandis*, the behaviors – of Twins (Oscars) as having the same causal powers, hence as taxonomically identical. (1987: 41)

If mental causal powers are mediated by neurological mechanisms, a difference in these local mechanisms would seem to dictate a difference in causal powers.

Fodor's argument can be summarized as follows:

- 1) Causal interactions are local (no action-at-a-distance).
- 2) So causal powers (in any robust, scientific sense) cannot differ without differences in local microstructure (/mechanisms).
- 3) But wide contents can differ without any difference in local microstructure.
- 4) So, wide contents can differ without differences in (robust) causal powers.

We may also conclude, given Fodor's methodological principle,

- 5) So, wide content individuation is not individuation by (robust) causal powers.

In short, according to Fodor the locality of causation makes mind/brain supervenience a necessary condition for a property to be a robust causal power; since wide contents don't supervene on the brain, they cannot be robust causal powers. The methodological principle establishes that Fodor's interest is in robust or genuine causal powers, not causal powers in any second-class or proleptic ("causal-explanatory but not genuinely causal") sense (see 1.4.3). It is not clear whether the sorts of local differences Fodor has in mind are longer-term structural differences or shorter-term dynamical differences. But either way, the upshot would still seem to be that if externalism is true there need be no

brain-state difference when content differs, and that this type of non-locality conflicts with the claim that causal powers should differ only if local properties do.

Burge's Response. Over several papers, Burge presents by far the most extensive and detailed response to Fodor. His initial response is to argue that this type of argument involves a “conflation” of causation – a relation between particular (token) events or states – with individuation – which is based on patterns (types) of causal relations.¹⁷⁰ The conflation abets drawing an inference (a *non sequitur*, to Burge) from a true claim about causation to a false claim about individuation: from the claim that there's a necessary dependence of token psychological causal relations on token neurological causal relations to the claim that psychological causal powers must be individuated in tandem with neurological causal powers. But neurophysiological individuation need not be sensitive to environmental differences, or to the same environmental differences, in the ways that psychological individuation is. So while token psychological causal relations require underlying token neurophysiological causal relations, psychological types need not supervene on neurophysiological types.

This argument, even if correct, doesn't go far enough, as Burge himself is (or becomes) aware. The issue is not just whether psychological types can cross-classify physical types – a claim Fodor obviously accepts – but whether this taxonomic difference marks a difference in robust causal powers. The challenge is to provide a reason for thinking that psychological types are robust causal powers given the claim that underlying physical types need not differ when the psychological types do, given what we know about physical causation.

¹⁷⁰ See Burge (1986: 16-17). I write “this type of argument” because his (1986) is obviously not a response to Fodor (1987); but the reply addresses the same type of argument and is part of his complete reply to Fodor, provided in Burge (1989) and to a lesser extent in his (1995).

Burge responds to this challenge by attacking the truth of premise (2) as a general principle that holds for all the sciences. Burge and Fodor agree that special science types will cross-classify tokens individuated in other ways in basic sciences. But Burge (1989: 310) claims, *pace* premise 2 in Fodor's argument, that the special sciences *also* cross-classify causal powers: "The special sciences track causal patterns that cut across one another." He writes:

Causation is local in the sense that each causal series of particular events between the environment and the individual's mental states necessarily "runs via", or presupposes the existence of, some causal relations (causal series of particular events) between the environment and the surfaces of the individual's body; and each causal relation between the individual's mental states and his behavior is necessarily associated with some causal series between the individual's neural states and his bodily motions. ... But individuation of mental and behavioral kinds, and the causal powers associated with those kinds, are non-local. (1989: 320 fn. 10).¹⁷¹

To support his claim of the non-local individuation of causal powers, he argues that causal-power individuation is relative to each science and its explanatory concerns. For example, a heart is typed by its normal function – pumping blood – in a normal bodily environment. Burge agrees that if a heart were transplanted into an organism whose physically identical counterpart pumped waste, it would have the same physical effects as the counterpart. But, he claims, the heart and the homologous waste-pump would not have the same causal powers "as typed by physiology." (1989: 312) Similarly, the conception of causal power in psychology is taken "not from some model drawn from the other sciences, but from the explanations that psychology provides." (1989: 316) So

¹⁷¹ As another way of putting the same point, he states (1989: 309) that "failures of supervenience of the mental on the physiological [at the type level] are compatible with psychology's assumption that gapless neurological chains must connect stimulus with response [at the token level] ..." (I've added the phrases in brackets for clarification.)

even if a difference in causal histories does not matter for individuating the causal powers of brains, it can, and does, matter for individuating psychological causal powers.

To defend the relativization of causal power individuation to each science, Burge argues that if causal powers were not relativized in this way, there would be no causal powers anywhere except in basic physics:

What is true is that physiological processes are *sine qua non* for psychological processes. But atomic and quantum processes are *sine qua non* for physiological processes. Few would advance the view that atomic or quantum processes are where the “real” causation goes on in physiology. Is there any strong ground for thinking that the “true nature” of causation involving psychological events is better revealed in physiology than in psychology? (1989: 318)

In short, his reply is that premise 2 of Fodor’s argument illegitimately generalizes to all sciences a way of individuating causal powers derived from the physical sciences. This way is (presumably) one in which externalistic differences do not make a difference to causal powers. Burge responds that despite the necessary dependence of token psychological causal chains on token neurological causal chains, the individuation of psychological types *and causal powers* depends on psychology’s explanatory concerns, not those of physical science. On this view, the properties that figure in psychological explanations are causal powers independently of their relation to the properties that figure in physical laws. Token psychological states enter into causal relations in virtue of their mental properties because these properties count as causal powers in psychology.

Comment on Burge’s Response. A relatively obvious, if minor, difficulty with Burge’s response is that it is too strong. The relativization of causal-power individuation (not just property individuation) to each science, if true, would guarantee that properties of any sort that appear in special science causal explanations (or in any “explanatory practice”) are causal powers independently of their relation to properties in lower-level

sciences.¹⁷² But it is not clear that the entire taxonomies of causal powers in the sciences are independent of each other. Many chemical reactions are explained in terms of interactions between basic physical particles. At least some specific perceptual capacities – edge detection, for example – may be identifiable with firing patterns in particular neurons or cell assemblies. If at least some special sciences, or parts of special sciences, allow for the reduction of the properties that play causal roles to those of lower-level sciences, Burge's across-the-board relativization of causal powers would be false.

This objection might be finessed by weakening the position to claim that at least some parts of the special sciences individuate causal powers independently of their relation to implementing properties, and that the parts of psychology that use wide contents in explanation would be among these. But even if this fallback position is defensible, I think Burge's response faces a separate, and fatal, objection anyway.

The serious objection is that Burge effectively ignores Fodor's deeper concerns about non-locality. His response tries to reconcile the necessary dependency of token psychological causal interactions on token physical mechanisms with the legitimacy of psychological causal claims – claims in which psychological types can play a serious causal-explanatory role. But he does not reconcile the locality of causation with the extrinsic nature of the resulting causal powers of intentional tokens. For he does nothing to explain *how* the causal powers of intentional tokens can differ with differences in their extrinsic properties on the assumption that the local physical mechanisms on which their token psychological causal interactions necessarily depend do not differ.

¹⁷² Burge holds that special-science properties supervene on lower-level properties; if they are relational properties, they globally supervene. My point is that his argument entails that these inter-level relations are irrelevant to determining the status of the supervenient property as a causal power; all that matters is its status within a science or explanatory practice.

In effect, Burge's response treats non-local and local properties alike in what is essentially a defense of *supervenient* causal powers. His argument for relativizing causal powers to each science rests on the claim that otherwise no supervenient properties would be causal powers. But this just means that wide contents count as causal powers because they belong to a special science, not because Burge has shown why their extrinsic nature is unproblematic. For suppose every special science type were narrow; Fodor's problem would disappear, and Burge's relativization of causal power individuation to each special science, however true, would have been completely irrelevant to resolving it.

As noted earlier, there *is* a problem about whether supervenient properties, wide or narrow, are genuinely causal (see 1.1).¹⁷³ But this is not the problem Fodor is worried about; after all, in the text in question he is arguing *for* the causal efficacy of *narrow* content, which is no less of a supervenient property. Moreover, the "few" who would advance the view that basic physics is where the "real" causation goes on are in fact the many reductive physicalists, who also do not distinguish between wide or narrow content when it comes to real causal powers (*neither* is). So without addressing Fodor's deeper concerns, Burge's argument invites a "proleptic" response from reductive physicalists, who would dismiss the psychological causal powers he defends as not robust, even if they are useful, and perhaps indispensable, in explanation.

This misdirectedness may arise from Burge's overlooking the significant role of physical microstructure in determining the causal powers of tokens. In Burge's defense, Fodor's rhetorical question – "How *could* differences in context affect the causal powers of one's mental states without affecting the states of one's brain?" – may well be

¹⁷³ In fact Baker (1995) quite appropriately provides a Burge-style response to Kim's Exclusion argument, which targets supervenient properties.

understood as a type-level question: How *could* contextual differences affect the classification of mental causal powers without affecting the classification of brain-state types? To which Burge answers: Easily! Each special science taxonomizes by its own causal powers; end of story.

But Fodor's challenge goes deeper: how *could* a token's causal powers differ if there is no difference in the microphysical structure of the token, however cross-classified that token might be at any level of explanation? On the assumption that the relevant physical properties of a token are determined by its constitution, Fodor's rhetorical question can be rephrased as follows: "How *could* special-science types taxonomize tokens *by robust causal powers* if the tokens don't differ *constitutionally* with each taxonomic difference?" Burge's response does not begin to address this worry, since he does not provide any alternative account of why these taxonomies should be considered *robustly causal* if the tokens remain *constitutionally* the same. Even if he is correct that extrinsic special science taxonomies are causal, he hasn't said a thing that would convince *Fodor* that they are.

In sum, Fodor and Burge end up talking past each other. On Fodor's view a property can play a causal-explanatory role in a science yet not count as a real causal power if it differs when local differences do not. For to accept that a token's causal powers can differ without a change in its physical constitution seems to force us to accept a way of instantiating token causal powers that violates the basic physical causal principles that operate in our world.¹⁷⁴ Thus, on his view only those classification schemes that satisfy this additional constraint can count as robustly causal. Burge clearly

¹⁷⁴ If we think of the local features in terms of implementing mechanisms or syntax, then the problem is that it seems that semantics cannot be implemented in syntax. This is Fodor's (1994) expression of the problem and the terms in which he offers a new solution. I discuss his (1994) below.

rejects Fodor's extra constraint for properties to count as real causal powers. For Burge, playing a causal-explanatory role in a science is sufficient for a property to count a genuine causal power; it is not also required that tokens must differ in microstructure whenever these properties differ. Presumably, on his view, if a property is really explanatory in a science, then it must be a real causal power.

I take it that the primary motivation behind Fodor's position is a commitment that follows naturally, although not logically, from intentional realism (see 1.3.2). Our best science seems to tell us that if there are mental states (tokens) that cause behavior and are semantically evaluable, then they are at least constituted in (if not identical to) brain states.¹⁷⁵ Given this view, Fodor's metaphysical principle tells us that wide contents can count as causal powers only if their instantiation by tokens involves a token brain-state difference (as well as standing in a relation).¹⁷⁶ (Advocates of the "extended mind" thesis, noted above, might disagree.) Perhaps it is not incidental that Burge's defense of mental causation has been accompanied by an increasingly non-committal attitude toward materialism, hence toward this corollary of intentional realism.¹⁷⁷

What makes Burge's position less convincing is that it lacks the support Burge seems to feel it has from other special sciences. The only reason it may seem that playing a causal-explanatory role is sufficient for a property to count as a causal power is because

¹⁷⁵ Vehicle externalists, or "extended mind" theorists such as Clark and Chalmers, would not agree with this statement. But such positions are controversial and, as noted in 1.2.3., I assume a more traditional view of the vehicles of content.

¹⁷⁶ Combined with his Modal Argument, the metaphysical principle would state more specifically that wide content differences make a difference to causal powers if their instantiations involve brain-state differences. If there is no difference in token brain states, then the causal powers of a token belief that water is wet and a token belief that twater is wet are the same – that is, they are determined by brain-state differences, and external relations play no role.

¹⁷⁷ This is most pronounced in his (1994), where he rejects the idea that a solution to the problem of mental causation requires us to embrace materialism. Kim (2000: 60-67) criticizes both Burge and Baker – rightly, I think – for offering a "free lunch" solution to the problem, in which we ignore metaphysical issues. Baker (1994) also seeks to "put explanation first" and avoid metaphysical stances – see 1.4.3.

analogous issues of local constitution do not arise for hearts, continents and many other special science entities. It is taken for granted that token hearts are constituted in (or identical to) bodily tissues, for example; “cardiac realism” takes local constitution for granted. This background assumption buys more than just the claims that hearts exist and are composed of muscle fibers. It brings along with it the possibility of reducing the causal powers of a heart to those of its muscle fibers and ultimately its physical particles. Burge dismisses this motivation too quickly when he states without elaboration that “few would advance the view that atomic or quantum processes are where the “real” causation goes on in physiology.” Avoiding explicit commitment to locally constituted intentional tokens does not make the central role of physical constitution in determining differences in causal powers go away.¹⁷⁸

But once that commitment is made, Fodor’s additional constraint cannot be simply ignored. And when it is not ignored, the problem arises immediately: how could the property of being a belief that water is wet *possibly* make a real causal difference – be a genuine causal power – if a token that has that property need not differ constitutionally from a token that has a distinct content property, or none at all? We need an excellent reason to think that differences in local constitution are merely sufficient for a difference in causal powers and not necessary. Burge may not agree that such a reason is needed, but Fodor probably speaks for many scientists as well as philosophers when he insists that it is.

In sum, whereas Fodor claims that causal laws that do not track local differences are “impossible” – that is, scientifically implausible – Burge replies that the special

¹⁷⁸ Pereboom’s (2002) defense of robust nonreductive mental causal powers, which denies that constitution is identity, is aimed precisely at blocking this inference from local constitution to a reductive explanation of intentional causal powers. See Chapter 4.

science laws don't follow this stricture. But he offers no explanation of *how* a token's content can make a real causal difference if its having that property, as opposed to another or none at all, need not involve a difference in its physical constitution. I take it that Burge offers no explanation because he thinks none is required. On his view it's enough for a property to be a causal power (and for its tokens to enter into singular causal relations in virtue of that property) if it plays the right role in explanation; if this is enough for biology and geography, then so too for psychology. This is a coherent position to take *if* it is true that that's enough for biology and geography – which is not at all clear. At the very least, this position requires ignoring the implicit but crucial role of local constitution in determining differences in real causal powers.

My criticism of Burge also applies to another point he makes about psychological explanation: that the sciences typically presuppose the individuation conditions of their types when the types are used in explanation.¹⁷⁹ His point is that wide psychological properties can be used in explanation even if their individuation conditions, including external relations, are never mentioned in those explanations or even if they are never explained in psychology.¹⁸⁰ This claim seems to imply that the external relations required for individuation are mere background conditions that do not affect the explanatory roles played by the types.

But Burge seems to be confusing two issues. It is quite right to distinguish the causal-explanatory role of a property from what constitutes that property. But it is another

¹⁷⁹ Burge (1989: 313-315) also used this claim to respond to Fodor's argument from the Formality Condition (see 2.2.)

¹⁸⁰ Burge's point is made in response to an objection to the effect that wide properties can't play explanatory roles in cognitive psychology because cognitive psychology does not make reference to the social relations that individuate them. He argues that the objection "conflates" the role of relations in individuation (which may be considerable) and in explanation (which may be nil).

issue to determine whether a causal-explanatory role is genuine or real, or instead proleptic or reducible. What constitutes a property is very much involved in this second issue. It's *because* external relations are constitutive of contents that they can't play *genuine* causal-explanatory roles – or so the opponents of semantic causal powers claim. So that's the claim that a defender of robust semantic causal efficacy has to face. Obviously, if a type *is* a robust or genuine causal power, then its tokens interact causally in virtue of that property; no disagreement there. But the Non-Locality objection arises right at this point. For Fodor, at least implicitly, no type can be a real causal power if it is instantiated by tokens in a way that makes no constitutional difference; Burge rejects this constraint. So even if Burge is right that the relations are in the background when the types are used in explanation, the relations make a difference when it comes to the question of whether this explanatory role is robustly causal or not.

My Response to Fodor. First, a preliminary remark: as noted above, Fodor intends to limit the scope of his objection to just some extrinsic properties (including semantic ones) by arguing that there are other extrinsic properties that are causal powers. So his intention is not to claim that there's something in the nature of all extrinsic properties that makes them causally irrelevant. It's just that that's the way things are for semantic properties and some other extrinsic properties, like being an H-particle, at least at worlds such as ours where causation is local. The remarks in which he tries to draw this distinction among extrinsic properties primarily involve the example of the property of being a planet, which on his view is both extrinsic and causal. I will argue in 3.4, however, that each of the reasons Fodor gives for making this distinction fail to show why being a planet is a causal power while being a belief that water is wet is not.

But even if we restrict Fodor's argument to just wide contents, I think it fails. I agree that on one view of what local microstructure is, his premises do lead to his conclusion. I think that view, in which local microstructure is intrinsic, is widely taken for granted, by Fodor as much as by anyone else in this debate. (I remind the reader of my reply to McGinn above, where I argued that 'local' is not the same as 'intrinsic': briefly, a property is *local* if it (or its microphysical base) is instantiated entirely within the spatial boundary of a concrete individual, while a property is *extrinsic* if an individual's having that property entails the existence of distinct individuals.¹⁸¹) But I also think there is alternative view according to which local microstructure is extrinsic, and that on this view Fodor's argument is invalid. In this case, local microstructure would be instantiated entirely in the brain, but it would be individuated extrinsically: two brain states that are constituted (locally!) of exactly the same type of neurons in exactly the same type of configuration could be instances of different types of brain states if the brains exist in relevantly different types of worlds. If so, Fodor's argument is only as good as the assumption about local microstructure that it presupposes.

I'll begin by restating Fodor's argument:

- 1) Causal interactions are local (no action-at-a-distance).
- 2) So causal powers (in any robust, scientific sense) cannot differ without differences in local microstructure (mechanisms).
- 3) But wide contents can differ without any difference in local microstructure.
- 4) So, wide contents can differ without differences in (robust) causal powers.
- and 5) So, wide content individuation is not individuation by (robust) causal powers.

¹⁸¹ Thus, the microphysical bases of extrinsic properties can be as "local" as the microphysical bases of intrinsic properties. What extrinsicness forbids is that an extrinsic property can be identified with, or strongly supervene on, an intrinsically individuated microphysical base. Also, while Fodor does assume that "local microstructure" – i.e. neural structure – is intrinsic, it doesn't follow that he also assumes that 'local' just means 'intrinsic'. However it wouldn't be surprising if he did.

Again, premise (1) is the claim about the locality of causal interactions that gets the ball rolling, and premise (2) embodies Fodor's "metaphysical" principle ("causal powers must supervene on local microstructure"). The type of modality here is physical necessity: premise (1) establishes a feature of causation in our world, not every metaphysically possible world; and the remaining premises must follow suit if the argument is valid. But by premise (3), wide contents violate premise (2): they do not supervene on local microstructure. Fodor cites Twin Earth as the reason for the failure of supervenience:

States of type X supervene on states of type Y iff there is no difference among X states without a corresponding difference among Y states. So, in particular, the psychological states of organisms supervene on their brain states iff their brains differ whenever their minds differ. Now, the point about Me and Twin-Me (and about Oscar and Oscar2) is that although we have different propositional attitudes, our brains are identical molecule-for-molecule; so it looks like it just follows that our attitudes don't supervene upon our brain states. (Fodor 1987: 30)

He then asks: "How *could* differences in context affect the causal powers of one's mental states without affecting the states of one's brain?" The implied answer is: they can't by (2), yet Twin Earth – which is assumed to be physically possible – appears to show that they can differ without any difference in brain states. If so, we can conclude, in (4), that contents can differ when causal powers do not, and that ordinary folk psychological states are not typed for robust causal-explanatory purposes.

It should be noted that in the above quote, Fodor appears to conflate the individuation of *molecular* structure with the individuation of *neural* structure, that is, molecular sameness or identity with neural sameness or identity. However, when presenting his metaphysical principle, Fodor explicitly explains "local microstructure" as local neural (or brain) structure, and throughout his argument his main emphasis is on mind/*brain* supervenience or identity. So a more considered interpretation is that "local

microstructure” should be understood unambiguously and explicitly as neural (or brain) structure in Fodor’s argument. I’ll use the phrase “neural structure” below to avoid confusion.

As noted, Burge attacked the truth of premise (2). I would add that the examples of chiral molecules in Ch. 2 provide strong empirical evidence that Fodor’s metaphysical principle is neither descriptively nor prescriptively true of even the basic sciences. If two molecules differ only in that one is left-handed and the other is right-handed, yet they have different causal powers, then it cannot be a general principle of science that causal powers cannot differ without a difference in local microstructure (of which neural structure is a particular case). Such molecules are identical in their constituents and configuration, yet their causal powers differ with their differences in spatial orientation. If the causal structure of the world turns out to be more fine-grained than differences in microstructure allow, then science will ignore Fodor’s metaphysical principle. *Pace* (2), causal powers in a robust scientific sense can differ without differences in local microstructure.

These scientific examples imply that (2) should probably be downgraded to a rule which has no deep normative implications even if it is generally followed in many sciences. It also implies that for (2) to be a threat to wide contents, it would have to be because psychological causal powers *in particular* cannot differ without a difference in local microstructure. However, for the sake of argument, I’ll accept that (2) applies to the case of psychological causal powers.

Also, premise (3) allows that extrinsic properties, and wide contents in particular, can have local components; being a parent entails being animate, and being animate will

make a difference to causal powers. And while the difference between being a water-thought and being a twater-thought is a matter of what's in the agent's environment, they have the same local components.

Now on what appears to be the standard view, neural structure is the *intrinsic* (and not merely local) component of wide content, in much the same way that husbands are adult males (intrinsic component) who stand in marital relations (external component), or T-particles are particles (intrinsic component) that stand in a relation to the faces of Fodor's dime (external component). If we assume this intrinsic view, then neural structure will supervene on molecular structure (on the assumption, which I grant for the sake of argument, that molecular structure is also intrinsic). Restricting our attention to psychology, Fodor's argument can be restated as follows, making the intrinsic view explicit in premise (2¹):

- 1) Causal interactions are local (no action-at-a-distance).
- 2) So causal powers (in any robust, scientific sense) cannot differ without differences in neural structure (mechanisms).
- 2¹) Neural structure cannot differ without differences in molecular structure.
- 3) But wide contents can differ without any difference in neural structure.
- 4) So, wide contents can differ without differences in (robust) causal powers.

(I omit (5) merely for brevity.) This argument is valid, and also sound if we assume the truth of (2¹) and agree that (3) is in fact supported by Twin Earth. Oddly, it may seem as if premise (2¹) is playing no role in his argument. But this is deceiving.

(3) is actually not supported by Twin Earth. At best Twin Earth shows that wide contents can differ without any difference in *molecular* structure.¹⁸² Oscar and Twin-Oscar are molecular Twins. The assumption that *therefore* they are neural Twins is

¹⁸² This is assuming that molecular structure itself is intrinsic – a claim that Twin Earth appears to take for granted, but which is not justified given the debate in chemistry about the nature of molecular structure. Still, for the sake of the current argument, I will grant this assumption.

question-begging if, *pace* (2^I), neural structure can differ when molecular structure does not. A plausible alternative to (2^I), defended below and in Ch. 4, holds that neural structure is the extrinsically individuated supervenience base of (or is identical to) wide content). There are thus two competing *hypotheses* about the nature of neural structure. Since Twin Earth shows only that Twins are molecular duplicates, we should rewrite premise (3) in a neutral way as (3*): wide contents can differ without differences in *molecular* structure. (3*) is supported by Twin-Earth but does not beg any questions regarding the individuation of neural structure. It may be accepted as true by advocates of (2^I) or of the extrinsic hypothesis, which I'll call (2^E).

The role of (2^I) can now be made clear. If we assume neural structure is intrinsic, and use (3*) instead of (3), we get the following as Fodor's argument:

- 1) Causal interactions are local (no action-at-a-distance).
- 2) So causal powers (in any robust, scientific sense) cannot differ without differences in neural structure (mechanisms).
- 2^I) Neural structure cannot differ without differences in molecular structure.
- 3*) But wide contents can differ without any difference in molecular structure.
- 4) So, wide contents can differ without differences in (robust) causal powers.

This is valid, and sound if (2^I) is true. However, if we assume that local microstructure is extrinsic, we get the following:

- 1) Causal interactions are local (no action-at-a-distance).
- 2) So causal powers (in any robust, scientific sense) cannot differ without differences in neural structure (mechanisms).
- 2^E) Neural structure can differ without any difference in molecular structure.
- 3*) But wide contents can differ without any difference in molecular structure.
- 4) So, wide contents can differ without differences in (robust) causal powers.

This argument is invalid. The truth of the premises does not support the conclusion. But the only difference between these arguments is (2^I) or (2^E). So Fodor's Non-Locality

argument stands or falls with the plausibility of the hypothesis that neural structure is intrinsic compared to the alternative that claims it is extrinsic.

I will make the case for the extrinsic hypothesis now before returning to my response to Fodor. Fodor himself argues for the intrinsic hypothesis mainly by calling the idea that we might distinguish between Oscar's and Twin-Oscar's brain states "mad" and "grotesque".¹⁸³ Usually, however, competing hypotheses are judged on their predictive success, explanatory power and theoretical fruitfulness, not on whether they are familiar or reflect pretheoretical intuition. The intrinsic hypothesis has not been defended on these grounds, if indeed it has been explicitly defended at all. On these grounds I argue that the extrinsic hypothesis fares better, or promises to fare better, than the intrinsic alternative.

The extrinsic hypothesis is that neural structure is the extrinsically individuated supervenience base of (or is identical to) wide content. So, for example, if *x* is a water-belief, there must be water in *x*'s world; by the extrinsic hypothesis, if *y* is the neural supervenience base of a water-belief, then *y*'s world also must not be otherwise empty, even if we cannot now specify what conditions must hold in *y*'s world. It is probable that the environmental differences that partly individuate wide content are closely related to the environmental differences relevant to individuating neural structure. In any case, there are several reasons for thinking this hypothesis is plausible.

First, cognitive neuroscience already individuates brain areas partly in virtue of the types of cognitive function they support.¹⁸⁴ (I assume "local microstructure" is or includes brain areas, parts of brain areas, or activity in brain areas.) Although physical differences and similarities play an essential role, brain areas are not individuated by

¹⁸³ See Fodor (1987: 34, 37-38).

¹⁸⁴ See Bechtel and Mundale (1999) for a critical review of individuation in neuroscience, as well as Banich (1994), Part I.

purely internal physical criteria. There is a taxonomically significant interrelationship between cognitive functions and brain areas at a neuroscientific level of description and explanation. The primary visual cortex, Broca's area and Wernicke's area, to mention some examples, are among the many cortical areas which are individuated in part by the specific types of information (visual, speech production and speech comprehension) that they process. So if at least some cognitive functions differ when the environment differs, then the extrinsic individuation of at least some brain structure is already part of neuroscience. Brain states, as essentially information-bearing states, are extrinsic to the extent that information is typed at least in part in terms of what is in the agent's environment.

It is also an open question what would be true of a Twin world where, for example, visual information is processed in a brain structure located where Broca's area is in the left frontal cortex (Brodmann's areas 44 and 45), while speech production processes occurred at the tip of the occipital lobe (Brodmann's area 17). Is their visual cortex in Brodmann's areas 44 and 45 and their Broca's area in Brodmann's area 17? Or does their visual cortex process speech and their Broca's area process visual information? The fact that such hypothetical questions may admit of no determinate answer tells against assuming a priori that neural structure is intrinsic.

Second, the possibility that cognitive neuroscience is just using a temporary method of individuation – one that will be useful until we have a better idea of what goes on at a lower, non-cognitive level – also does not undermine the extrinsic hypothesis. For there is no assurance that non-cognitive levels will be intrinsically individuated. As I will discuss at length in Ch. 4, some quantum chemists argue that molecular structure is an

extrinsic property. (In their terms, it is not context-invariant.) They argue that the structural descriptions of molecules by which they are classified into distinct chemical natural kinds are only true of them when the molecules are embedded in, and interacting with, a particular kind of environment. Otherwise there is no physically justified way to apply these descriptions to molecules, on the assumption that quantum mechanics has superseded classical mechanics. If a molecule has a true structural description, it follows, at the very least, that its world is not otherwise empty. Otherwise the molecule is an interacting bundle of particles which cannot in general be typed as a token of a particular chemical species.

One consequence of this is that even though Oscar and Twin-Oscar are molecular duplicates by hypothesis (and I accepted them as such), chemistry does not require that they be individuated that way. So while premise (3*) in the above arguments is true on the assumption that molecular structure is intrinsic, this assumption is not justified by contemporary chemistry. It is also worth noting that the extrinsic individuation of brain structure does not entail that brain states are not located in the brain, any more than the extrinsic nature of molecular structure means that molecules are not located where their atomic constituents are. The “extended mind” thesis discussed in Ch. 3 is compatible with the extrinsic hypothesis but does not follow from it.

Third, since it is now being challenged, the intrinsic hypothesis must motivate the claim that neural structure is, will, or ought to be intrinsically individuated. But the empirical fact of neural plasticity motivates wide individuation of neural structure, in the following way.

Neural plasticity is the fact that the neural (synaptic) connections by means of which our brains process information undergo both structural (long term) and dynamical (very short term) changes in response to experience through our lives.¹⁸⁵ Changes in brain structures (from a token of N_1 to a token of N_2 , for example) are not the same thing as taxonomic differences in brain structures (e.g. the difference between N_1 and N_2). A change from an N_1 token to an N_2 token *presupposes* a taxonomic scheme which distinguishes N_1 tokens from N_2 tokens. But pre-taxonomic observable changes are the starting point for a scientific taxonomy of explanatory similarities and differences; the taxonomic trick is to determine which changes are explanatorily significant. Thus, the myriad, constant changes in synaptic connections and so forth will count taxonomically – as changes from one microstructural type to another type – once we determine which changes are explanatorily significant. Moreover, the significant changes must potentially occur in all the physically distinct and neurally plastic brains that we want to include in our generalizations.

Neural plasticity makes it very likely that explanatorily significant changes in synaptic connections will be picked out partly in terms of the observable changes in cognitive function – changes which presuppose a cognitive-functional taxonomy – that occur as a result. For reasons of both practicality and explanatory generality, differences in cognitive-functional types will help guide the typing of changes in the neurally plastic brain. And to the extent that differences in cognitive function are extrinsic – dependent in part, for example, on behavioral differences – the typing of microstructure in the neurally plastic brain will also be extrinsic. If Oscar's water-thoughts are about water partly in

¹⁸⁵ See Churchland (2002) for a general discussion of neural plasticity; Bates (1994/2001: 139) notes that neural plasticity can differ depending on the cognitive function subserved, and that the brain systems that support languages “show an extraordinary and perhaps unprecedented degree of neural plasticity.”

virtue of the fact that water has caused (perhaps via a long, indirect chain) his water-thoughts, and if Oscar's neural structure is also causally related (perhaps via a long, indirect chain) to water, and if Oscar's neural structure *changes* partly as a result of these causal interactions with the environment, then these changes are very likely to be individuated for explanatory purposes partly in virtue of the different types of thoughts and behavior that Oscar exhibits as a result. If it is fair to assume that the developmental trajectories of brains and minds (or cognitive capacities) are importantly related, it also seems fair to assume that this complex method of taxonomizing brain states by both cognitive and physical factors will continue.

An intrinsic taxonomy, in contrast, would not be able to depend essentially on ordinary cognitive-functional differences to distinguish significant neural changes from insignificant ones. This will make a purely intrinsic taxonomy practically unattainable and probably lacking in generality. No two brains are ever physically exactly alike, after all. But still, if our brains had been such that they underwent no or hardly any changes, perhaps we might have been able to get away with using purely internal criteria to individuate neural structure. There is no obvious need to appeal to cognitive-functional differences to individuate unchanging, virtually rigid neural structures. But this empirical possibility is not actual; cognitively significant synaptic connections in the brain are not rigid. This makes it very difficult to motivate a purely intrinsic taxonomy of the brain states that underlie cognition. Again, the intuitive fallback claim – that the neurons are all in the head, so we can in principle assume that neural structure is reducible to molecular chemistry – is not a fallback position at all. The reduction may be true without

undermining the extrinsic hypothesis at all, given the molecular structure debate in chemistry.

Which of these two competing hypotheses about neural structure will turn out to be the right one also cannot be decided on an *a priori* basis. This is important for the following reason. I am not arguing that the extrinsic hypothesis is true. I am arguing that it is highly plausible. It appears to be true of actual neuroscientific taxonomy, and it is appears likely that this complex method of individuation will continue. This is sufficient to show that the extrinsic hypothesis *might* be true, which in turn suffices to show that Fodor's Non-Locality argument does not succeed in ruling out wide contents as causal powers. For his argument depends essentially on an opposing hypothesis whose truth can no longer be taken for granted. All we can do ahead of empirical study is to draw out the consequences of each hypothesis, particularly in relation to the purposes of causal explanation, and to see how they fare in relation to what we already know. For my money, a hypothesis that promises to demystify intentional causation is better in at least one important dimension than one that forces us to deny that the contents of our thoughts make any difference to what we do.

There is a strong intuitive objection to my counterargument that is likely to be raised. If I allow that Twins are physically possible, the objection goes, then if I am right about causal-power taxonomies, we would be classifying Twin brains as having distinct causal powers. This is not merely "grotesque", as Fodor might say. Surely science will want to generalize over Oscar and Twin-Oscar. For surely our causal laws ought to cover molecularly identical states in all physically possible worlds, and the reason they ought to is that molecularly identical states behave the same! Moreover, it would violate a central

causal-explanatory aim of science – to obtain general laws – if we classified Twins under different causal laws. Briefly put:

- 1) Molecularly identical states behave the same.
- 2) So, science ought to count molecularly identical states as the same causal powers.
- 3) Wide content individuation allows that psychological states with molecularly identical supervenience bases can be classified as distinct causal powers.
- 4) So wide content individuation will not classify psychological states in a way that is scientifically justified.

This argument is similar in spirit to Stich's Replacement argument of Chapter 2, in which he claims we ought to explain behavior under autonomous (local) descriptions because otherwise we will miss generalizations that ought to cover Stich and his Replica. Wide individuation, Stich claims, "slices the pie too finely."

Note that this new argument would go through even if it were true that neural structure is extrinsic. For the problem it raises is that intuitively we expect that Oscar's and Twin-Oscar's physical states will have the same effects. This problem can be generalized. Given an identity relation between local physical properties of different agents, is there any scientifically justified reason for classifying their causal interactions under distinct causal laws, as wide content individuation would allow? If they behave the same, as we would expect, isn't it of overriding scientific importance to take account of this fact in our causal laws by classifying identical molecular structures as the same causal powers?

To counter this argument, it is enough to point out that (1) is false and that as a result (2) is unjustified as a general claim about science. Once again, chiral molecules provide a counterexample to premise (1). Mirror image molecules differ only in their orientation in space, and this extrinsic difference makes a causal difference. So (2) is not

descriptively true of science, nor is it prescriptively true, even if it is a habitual rule of thumb. There is ample scientific justification for individuating causal powers in a more fine-grained way than that determined by identity of molecular structure. This fineness of grain is motivated by experimentally confirmed differences in the causal powers of extrinsic properties. Science *does* seek to generalize, as Stich and many others often proclaim, but it also seeks to avoid *overgeneralizing*. *Pace* Stich, wide content individuation does not “slice the pie too finely” if in fact the world behaves that way.

Moreover, there is no scientific basis for the conclusion that Oscar and Twin-Oscar will behave the same, given contemporary chemistry. This can be shown with the following thought-experiment. Imagine a universe with just one α -methyldopa molecule in it; the universe might be ours with its current contents imagined to have disappeared. Surely, one might claim, the causal powers of that molecule will not differ from those it has in our crowded universe. Why should it matter to its causal powers if there is nothing else in its universe? But wait. If the molecule is truly isolated, then it has no determinate structure at all, if quantum mechanics is correct. We have no justification for labelling it an α -methyldopa molecule. There's not enough stuff in the universe to determine its identity, hence its causal powers. If our best science is correct, the scenario we are imagining is either physically impossible or crucially underdescribed.

So then let us imagine this molecule in a chiral environment – for example, by adding a carrier molecule that is enantioselective for the left-handed form. We will then know what we have. But this only means that its orientation matters for individuating its causal powers, since by hypothesis the carrier is enantioselective. If it is the left-handed enantiomer, we know it will have the desired antihypertensive effect in the appropriate

circumstances; if it is the right-handed enantiomer, it would be inactive (or at least would not have the desired causal power).

In short, if you imagine a scenario where you have one molecule in an otherwise empty universe, it may be indeterminate what its identity and causal powers are. Yet if we add what we need to determine its identity, hence what causal powers it has, the causal laws it falls under may well be distinct from those it would have fallen under had it been the antipode.

Twin cases need be no different. We have no scientific justification for the conclusion that Twins *should* be covered by the same causal laws, for the information that they have the same molecular structure is insufficient to determine this. If we add the information that Twin-Oscar is a mirror image of Oscar, we would conclude that they should *not* be covered by the same causal laws. If we add the information that they are identical in the way two left-handed α -methyldopa molecules are identical, then we would conclude that they'll have the same causal powers, but this will depend on their spatial orientation. Thus, since our best chemistry does not rule out laws that distinguish mirror image molecules for causal explanatory purposes, we cannot claim *a priori* that content properties are not causal powers due to a general scientific demand for explanatory generality and the admitted physical possibility of Twins.

Whether Fodor would put forward this objection is an open question, but it is likely that he would. In his (1994) attempt to guarantee content a causal-explanatory role, he *assumes* our causal laws cover both Oscar and Twin-Oscar due to the sameness of their local microstructure ("syntax"), for his *goal* is to get content to covary with "syntax". Moreover, he reaches his goal in part by arguing that Twin concepts are

nomologically impossible by our best chemistry, because anything that's as much like H₂O as XYZ *just is* water. Informational semantics will assign the same contents to Oscar's and Twin-Oscar's thoughts.¹⁸⁶ So on Fodor's solution, the primacy of local microstructural identity in determining sameness and difference of causal powers is preserved, as is the assumption that neural structure is intrinsic.

Whether one thinks this solution constitutes a good defense of wide semantic causal powers or not (I don't think it does), Fodor's appeal to chemistry in fact leaves the critical questions open. H₂O is not a chiral molecule, but the principle is the same. Even if Fodor is right, Oscar and Twin-Oscar may fall under a law that recognizes their extrinsic properties. For besides knowing that they share the same molecular structure, if they also fall under the same causal law we can conclude that they're *not* mirror images. This would allow that the causal powers they share are *not* determined locally even if they are molecularly identical. Fodor *assumes* that they are; on his view, only "syntax" is genuinely causal because only "syntax" supervenes on neural structure. But this may be false – as I will explicitly argue in Ch. 4.

3.2.2. Dretske's Argument from (Temporal) Non-Locality

Argument. Dretske (1995) presents, and then replies to, a closely related version of Fodor's Non-Locality objection that focuses what might be called the temporal non-locality of wide contents. This is the fact that at least some of the relations individuating

¹⁸⁶ In effect, what's nomologically impossible is that anyone could have wide thoughts about conceptually possible chemical kinds. *Wide* thoughts about conceptually possible chemical kinds are nomologically impossible, given informational semantics and chemistry. (We can *imagine* XYZ; but our account of the contents of such thoughts need not be informational. See my response to Fodor and Pylyshyn in 3.3.2). The kinds that Burge emphasizes – "contract", "arthritis" and so on – are not as easily handled as these cases of scientific kinds, but this would not impugn Fodor's argument.

the contents of current mental states involve entities that existed in the past (relative to the time of the occurrent thought) as *relata*. For example, the content of my current belief about Socrates is individuated partly in terms of some causal-historical relation between my mental state and the long-dead Socrates; *mutatis mutandis* for my current belief about Socrates' death, a long-ago event. I do not think Dretske's argument provides any substantially different elements from Fodor's, and can be assimilated to Fodor's. But it has been widely discussed in the literature and so should at least be presented here.

Dretske begins by noting that there is no *general* metaphysical problem with citing temporally remote events as causes of current behavior. If Jimmy stutters because he was dropped on his head as an infant, then that past event is a temporally remote cause of the neurological condition that now operates as the proximal cause of his stutter. If a brick's falling on Johnny's head leads to the same neurological condition and stuttering, it just means that Johnny and Jimmy have different remote causes of the same effects. So there is no difficulty in supposing that historical events can be causes of current behavior: "Remote causes *are* causes even though their effects are brought about via more proximal events." (1995: 153)

The problem with contents, he argues, is that while they involve historical relations they are not cited in psychological explanations as *remote* causes of current behavior. If I pick up a glass of water now because I had a desire to drink a moment ago, the content of my desire is supposed to be an immediate or proximal cause of my picking up the glass. In Dretske's words:

The point is rather that [externalism] somehow makes the *here-and-now* (i.e., proximal) explanatory relevance of experience and thought into a *there-and-then* causal relevance. This is objectionable because ... beliefs, desires, and feelings – the *reasons* for which we act – have an "immediate" causal/explanatory relevance

to our intended behavior. By making the causal/explanatory relevance of the mind (temporally) remote, [externalism] denies this basic intuition. (1995: 154)¹⁸⁷

The difficulty might be clarified as follows. Many of the relations and relata that help individuate the contents of our current thoughts occurred long ago: a perceptual relation to Socrates, for example. For the sake of argument, we can also assume that there is an unbroken, if complex, causal chain from those past events to current states – e.g. a chain winding through generations and societies that links my current Socrates-beliefs with past perceptual relations to and beliefs about Socrates. But contents are cited as proximal causes of current behavior; they're not remote causes, like Jimmy's fall. And the problem with that – which Dretske doesn't make explicit – is that these here-and-now instantiations of different beliefs involve the *same vehicles*; to borrow Stich's phrase, the historically-based differences in individuation do not "leave their trace" in the form of differences in here-and-now physical states.¹⁸⁸ This problem would also afflict the so-called remote causes of Jimmy's and Johnny's neurological conditions *if* these neurological conditions were vehicles with properties representing a fall or a striking brick: we would doubt the proximal causal relevance of these representational properties given the hypothesized sameness of vehicle. In short, if the historical, there-and-then differences that matter for individuating vehicles make no difference in the here-and-now properties of those vehicles, it is hard to see how the contents of our current thoughts can make a causal difference.¹⁸⁹

¹⁸⁷ In this quote I have substituted "externalism" where Dretske writes "the Representational Thesis", which asserts a form of informational, hence externalist, semantics (see 1.7.3).

¹⁸⁸ Elsewhere, at (1995: 186 fn. 17) Dretske does acknowledge a distinction between "the here-and-now character of the belief (what I, in chapter 1, called the belief *vehicle*) and the here-and-now character of *what* is believed (the belief-*content*). It's the fact that the vehicles are the same for different contents that creates the problem.

¹⁸⁹ Perhaps the local difference in the "temporal" Non-Locality problem is not in the vehicle as such but in the time that it is tokened? This is possible, although we can build simultaneity into the case to get identical

This version of the Non-Locality objection might be summarized as follows:

- 1) The contents of current mental states differ with differences in historical relations.
- 2) The vehicles of current mental states need not differ with historical differences.
- 3) So differences in content due to different historical relations need not make a difference in current vehicular properties.
- 4) But causal powers are temporally local: only current properties (or properties that make a difference to current properties) can make a causal difference.
- 5) So contents are not temporally local, and so are not causes.

Because the problem stems from the sameness of here-and-now properties of vehicles, the core difficulty is just that of Fodor's Non-Locality argument (or the Ur-Argument). In sum, differences in content can stem from extrinsic relations, including relations to spatially or temporally non-local entities. But when these (there-and-then) differences which distinguish tokens are not reflected in differences in the spatially and temporally local (here-and-now) properties of those tokens, the claim that contents make a causal difference conflicts with our causal intuitions. I have already argued how this conflict is illusory; Dretske does not provide any new reasons to think that it isn't. But for the sake of thoroughness I will rehearse Dretske's reply and provide a short response.

Dretske's reply. Dretske responds that while the vehicles of content are here-and-now states that causally explain movement, their contents explain *behavior*, and behavior is not identical to movement. In his terms, contents are structuring causes of behavior, not triggering causes of movement. A structuring cause explains why a certain type of process exists; a triggering cause is what sets off a token of that process now. Historical facts, which are temporally extrinsic, play a role in structuring-cause explanations, not in triggering-cause explanations. So contents are not deprived of a causal-explanatory role

times of tokening. The phrase "same neurological condition" (or "same vehicle") leaves this open, although I doubt Dretske has a local difference of this sort in mind.

by their being temporally extrinsic because the type of explanatory role they play essentially depends on their temporal extrinsicness.¹⁹⁰

For example, the reason why seasonal changes cause changes in plant color may be that the difference in plant color historically helped camouflage ancestors of today's plants from a rapacious insect. Causal-historical relations, such as those involved in natural selection, don't explain the occurrence now of a token series of events that runs from rising temperature to chemical activity to color change. They explain why that type of process exists rather than another. As Dretske (1995: 156) puts it, the causal pattern is:

Natural selection $\rightarrow_s$ [Warm Days $\rightarrow_t$ Chemical Activity $\rightarrow_t$ Color Change],
where the brackets frame a process and the arrows (which I subscript to distinguish "structuring" from "triggering") represent causal connections. Natural selection is the structuring cause of the process, not the triggering cause of warm days; but warm days are the triggering causes of chemical activity. The chain of events that compose the process is a physically describable and explainable causal series; but the process as a whole has an explanation that involves essential reference to historical facts.

Similarly, contents explain why a given type of internal state is wired up to a given type of movement; the whole internal-state-to-movement process *is* the behavior, and the historical facts that individuate content explain why that process-type exists rather than another. In the case of content, learning history – in particular, associative learning – takes the place of natural selection. The causal pattern is like this:

Learning history $\rightarrow_s$ [Internal state $\rightarrow_t$ Bodily Movement].

¹⁹⁰ Dretske (1988) is his full account of behavioral processes and structuring causes; a summarized version is provided in his (1995).

Each causal chain from internal state to movement is a neurophysiological matter; content is not a temporally remote member of that causal chain. But the historical facts explain why that process, that internal structure, exists rather than another. Dretske (1994: 186 fn 20) adds that structuring causes explain current behavior, not dispositions, because they “explain each and every tokening of the process they structure. In explaining why the dog salivates when the bell rings, I explain why it salivates each time the bell rings.” But this means only that each token of a given process-type can be given the same structuring-cause explanation. It does not mean that content is a triggering cause of current behavior.

My response. Dretske’s accounts of behavior and structuring causes have spawned a cottage industry of criticism that need not be reviewed here.¹⁹¹ However, one serious criticism suffices to show why Dretske’s response is unsatisfactory for anyone (such as myself) who is seeking a robust causal role for content. The objection, voiced by Block (1990) and Cummins (1988) among others, is that however plausible Dretske’s account of structuring causes may be, it does not give us what we wanted: a role for contents as triggering causes. We wanted an explanation of how the content of Clyde’s belief that there’s beer in the fridge can be the cause of his getting off the couch now; that is, we wanted a token’s having the content it does to make a causal difference, such that if that token had had a different content, it would have had a different effect. The fact that Dretske finds another explanatory role for content is no consolation.

¹⁹¹ See McLaughlin (1991), Baker (1995a: 56-62), Block (1990: 153-154), Cummins (1988), to name a few.

3.3.3. Summary of Non-Locality Arguments

McGinn's Non-Locality argument claims that relational properties can't be causes because they are not "locally" instantiated; like relations – since they are relations – they are "spread out" in the world and not contained in a single spatiotemporal location. I argued that McGinn's argument fails because it rests on confusing what's "constitutive of" or essential to an extrinsic type, and what the "constituents of" a token of an extrinsic type must be. I also argued that the specific "referential" problems raised by Fodor and Pylyshyn only raise problems for the naturalization of content, not for the causal efficacy of content.

Fodor's Non-Locality argument claims that semantic properties can't be causal because there can be differences in content without differences in local properties, and there can be no causal difference without a local difference. Dretske provides a variation on this theme. I argued that Burge's reply to Fodor and Dretske's response to the problem are not adequate. Even if Burge is correct that causal powers need not supervene on local microstructure, he does not show why such individuation schemes are causal in a way that would satisfy Fodor. Dretske's response, in which contents are structuring causes, doesn't satisfy our original demand for a way to show how they are (what he would call) triggering causes.

In my response, I argued that Fodor's argument stands or falls with the position one takes regarding the nature of local microstructure – that is, brain states. On the intrinsic hypothesis, it is valid; on the extrinsic hypothesis, it is not sound. I also provided reasons for thinking that the extrinsic hypothesis is plausible.

3.4. ONE LAST ARGUMENT, AND ON BEING A PLANET

I will wrap up this chapter with a brief argument from Fodor that does not fit into any other category, and a discussion of Fodor's many defenses of his claim that the relational property of being a planet is a causal power even if wide contents are not. The latter discussion will also serve as a review of the arguments of the previous chapter.

3.4.1. Fodor's "Technical" Argument

Fodor raises what he calls a "technical" reason for why contents can't be causes. This argument does not target wide contents per se, but it is one of the reasons he gives for his "metaphysical prejudice" against intentional causation. His argument goes as follows:

If thoughts have their causal roles in virtue of their contents per se, then two thoughts with identical contents ought to be identical in their causal roles. And we know that this is wrong; we know that causal roles *slice things thinner* than contents do. The thought that $\sim \sim P$, for example, has the same content as the thought that P on any notion of content that I can imagine defending; but the effects of entertaining these thoughts are nevertheless not guaranteed to be the same. Take the mental life in which the thought that $P \ \& \ (P \rightarrow Q)$ immediately and spontaneously gives rise to the thought that Q ; there is *no* guarantee that the thought that $\sim \sim P \ \& \ (P \rightarrow Q)$ immediately and spontaneously gives rise to the thought that Q in that mental life. (1987: 140)

In short, logically equivalent contents should count as the same contents, but logically equivalent contents may well have distinct causal roles.

Response. The argument equivocates between the individuation of content in an ideal rational agent, for whom logically equivalent contents are also *effectively* equivalent, and the individuation of content for ordinary agents, for whom logically equivalent contents need not be *effectively* equivalent. The case is not unlike that of someone who thinks that Superman is bald and that Clark Kent is not bald. We introduce

some form of mode of presentation – perhaps a sense, perhaps a causal or computational role – to account for the fact that while the concepts “Superman” and “Clark Kent” have the same content qua reference, their content is not exhausted by their reference, and they may therefore play distinct causal roles in an imperfect knower’s mental economy. Just so, someone who thinks that $P \ \& \ (P \rightarrow Q)$ entails Q and that $\sim \sim P \ \& \ (P \rightarrow Q)$ doesn’t entail Q is just an imperfectly rational agent. Presumably a similar explanation might be given for the fact that while P and $\sim \sim P$ have the same content as a matter of logic, logical equivalence need not exhaust the content of these concepts, so they may play distinct causal roles in that agent’s mental economy. “Superman” and “Clark Kent” will always have exactly the same meaning to a perfect knower, and “ P ” and “ $\sim \sim P$ ” will always have the same meaning for a perfectly rational agent. But we’re not perfect.

3.4.2. On Being a Planet

In the course of presenting his various arguments against wide intentional causal efficacy, Fodor repeatedly emphasizes that the relational property of being a planet is a causal power. (He later adds being a meteor.) His claim is that there is nothing wrong in general with relational properties being cited as causes; it’s just wide contents that are flawed. His argument amounts to a proof by cases, in which he argues that the property of being a planet, unlike (for example) the property of being a WATER-concept, survives each argument that he presents. In what follows I will look at each of his cases to see if the properties of being a planet and (e.g.) being a WATER-concept do in fact separate for the reasons he gives, such that one counts as a causal power and the other does not. Since Fodor’s remarks are relatively brief, I’ll quote them in full and respond to each in turn.

(1) Fodor's initial remarks are made in the course of his Non-Locality argument:

Thus, 'being a planet' is a relational property par excellence, but it's one that individualism [taxonomy by causal powers] permits to operate in astronomical taxonomy. For whether you are a planet affects your trajectory, and your trajectory determines what you can bump into; so whether you're a planet affects your causal powers, which is all the individualism asks for. Equivalently, the property of being a planet is taxonomic because there are causal laws that things satisfy in virtue of being planets. By contrast, the property of living in a world in which there is XYZ in the puddles is *not* taxonomic because there are *no* causal laws that things satisfy in virtue of having *that* property. ... The operative consideration is, of course, that where there are no causal laws about a property, having the property – or failing to have it – has no effect on causal powers. (1987: 43)

Response. If "being a planet" survives Fodor's Non-Locality argument, this quote does not show how or why. For his claim *here* is that if a property appears in a causal law, then it is a causal power; since being a planet appears in astronomical laws, it is a causal power. But if *that* is sufficient for a relational property to be a causal power – as Burge or Baker have argued (see 1.4.3) – then if there are psychological laws in which wide contents figure, then wide contents are causal powers too, and the two cases do not separate. But Fodor accepts that there are psychological laws; in fact, his commitment to psychological laws motivates his intentional realism.¹⁹² So being a planet and being a WATER-concept do not separate.

Another point, noted by Burge (1989: 321 fn. 12), is that the property Fodor cites in contrast with being a planet – the property of living in a world with XYZ – isn't a content property, nor does it have anything directly to do with having a given type of content. Living in an XYZ-world is necessary but far from sufficient for having XYZ-thoughts. So Fodor's use of this example of a content property that couldn't be a causal power is purely rhetorical.

¹⁹² See Fodor (1990: 178-85), also cited in 1.4.2, footnote 40.

(2) Further remarks are made in the context of his Modal Argument:

Compare cases of relational properties that really are causal powers, like being a planet. Being a planet is a causal power in virtue of, for example, its contingent (a fortiori, nonconceptual) connection with having a Keplerian orbit. That is, being a planet is a causal power because it is true and contingent that, if you have molecularly identical chunks of rock, one of which is a planet and the other of which is not, then, *ceteris paribus*, the one which is a planet will have a Keplerian orbit, and *ceteris paribus*, the one which is not a planet will not. (1991: 19)

He continues in the same vein regarding the property of being a meteor:

Take a pair of rock-twins such that one is a meteor and one is not. Then (*ceteris paribus*) craters will be among the effects of the first but not among the effects of the second. And the relation between the difference between the rock-twins and the difference between their effects is nonconceptual. So it is all right for being a meteor (rather than a meteor twin) to be a causal power in virtue of the fact that meteors cause craters and meteor twins do not. (1991: 20)

He adds that the ability to produce *meteor* craters is not a causal power over and above the ability to produce craters:

[I]t is conceptually necessary that, whereas a crater caused (in the right way) by a meteor is a meteor crater, a crater caused (in whatever way) by a molecularly identical nonmeteor is not. Analogously, its ability to cause sunburns is not a causal power that the sun has over and above its power to produce burns. (1991: 20 fn. 17)

Response. In the Modal Argument (see 2.4.2), Fodor he argues that only if a difference in effects is not conceptually connected to a difference in causes does the difference in causes count as a causal-power difference. Wide contents allegedly fail this test because the difference between water-behavior and twater-behavior is conceptually connected to the difference between water-thoughts and twater-thoughts. So these are the same causal powers and the relations by which they differ are causally irrelevant.

Fodor's final remark can be ignored. Not only does it contradict the previous ones – it allows that meteor-twins can cause craters – but it is not objectionable to anyone. My

power to cause Carrie-actions is not distinct from my power to cause actions, and my Twin does not have the power to cause Carrie-actions. So I will focus on the first two.

Fodor claims that planets and molecularly identical chunks of rock that are not planets will have distinct causal powers. A planet's property of being a planet is a causal power in virtue of there being a contingent relation between its being a planet and its having an elliptical (Keplerian) orbit; presumably it is physically possible that its orbit might have been circular. A molecularly identical piece of rock will not have an elliptical orbit. So the difference in effects is nonconceptually related to the difference in causes.

However, this is misleading. A molecularly identical chunk of rock will not have an elliptical orbit because if it is not a planet it will not have any orbit at all, since being a planet is a relational property which "involves" the relation of orbiting a star. In other words, that a planet's orbit is *elliptical* is contingent, but that anything that is a planet must orbit a star in some fashion or other is noncontingent. So the difference between the planet and the mere chunk of rock (one orbits a star, the other does not) is noncontingently related to the difference in their effects (one has an orbit, the other does not). So being a planet is not a causal power by Fodor's Modal Argument. So being a planet and being a WATER-concept will not separate.

The case with meteors and meteor-twins is harder to assess. If anything that is a meteor is something that causes craters, then the difference between a meteor and a meteor-twin is nonconceptually connected to the difference in effects. If not, then I see no reason to think that a meteor and a meteor-twin would not crash into the Earth in exactly the same way given the same trajectories. That is, their causal powers would be exactly

the same by Fodor's Cross-Context test, and so being a meteor would fail to be a causal power (just as being a WATER-concept does). Again, the properties do not separate.

Being a planet would also seem to fail the Cross-Context test, since a molecular twin of Earth, if placed where Earth is, presumably would have the exact same orbit. Martin Davies has made similar observations – that “it cannot be the case *both* that a planet has characteristic causal powers and not merely those of a physically similar chunk of matter that is not a planet, *and* that causal powers have to be compared across contexts or environments quite generally” – as Fodor (1989: 158 fn. 9) is well aware. This leads to the last set of remarks that I will discuss.

(3) Fodor makes these remarks in the course of the Cross-Context argument:

[Davies's objection] confuses the question whether *being a planet* is taxonomic (which it is; two things that differ in that property ipso facto differ in their effects in many contexts) with the question whether *being this piece of rock* is taxonomic (which it isn't; two things that differ in *that* property do not thereby differ in their effects in *any* context). Once again: a difference between properties P and P1 can affect causal powers (can be taxonomic) only when there is a situation S such that the instantiation of P in S has, ipso facto, different effects from the instantiation of P1 in S. By this criterion, the difference between *being a planet* and *not being a planet* affects causal powers because there are situations in which something that's a planet has, ipso facto, different effects from something that isn't. By contrast, the difference (in content) between the thought that water is wet and the thought that water₂ is wet does *not* affect the causal powers of tokens of these thoughts: there are no situations in which one thought has, ipso facto, different effects than the other. ... (Compare a real – taxonomically relevant – difference in content; e.g. the difference between my thought that water is toxic and your thought that it's potable. Tokens of these thoughts differ in their consequences in all sorts of situations.) (1989: 158 fn. 9)

Response. To the extent that I understand this remark (and I admit to some confusion about exactly what is meant), Fodor doesn't seem to be giving a *reason* why being a planet is a distinct causal power from being this piece of rock. The question is whether an item that satisfies both these properties has causal powers in virtue of its

being a planet that are not those it has in virtue of its being a piece of rock; equivalently, it is whether an item that satisfies both properties has different causal powers from an item that satisfies just the property of being a piece of rock. But Fodor compares being a planet with not being a planet – which would include (e.g.) being a pigeon. If that is the response, then being a WATER-thought will clearly differ in causal powers in many contexts from non-WATER-thoughts – e.g. being a pigeon. The same goes for being this piece of rock and not being this piece of rock (e.g. being a pigeon). So as far as I can tell, the argument doesn't address the question; it amounts to pounding the table.

In summary, in none of these cases has Fodor successfully distinguished between one relational property that is a causal power – being a planet – and another that is not – e.g. being a WATER-thought. So he has provided no reason that would undermine my claim that the intuitions driving much of the problem of externalism arise from a general metaphysics of relational properties that applies to both properties.

3.5. CHAPTER SUMMARY

In this chapter I presented versions of what I consider to be the critical objections to wide intentional causes, the Non-Locality objections. I divided these into two basic sorts, “spatiotemporal” and “referential” versions. Fodor's “spatial” Non-Locality argument relies on the claim that there can be no causal difference without a “local” difference. McGinn's argument relies on the claim that semantic properties, as relational properties, are just relations and so aren't instantiated “locally”. I showed that McGinn's argument is unsound; it relies on a view of what relational properties are and how they are instantiated that is controversial and undefended. I argued that Fodor's argument

depends essentially on the hypothesis that local microstructure is intrinsic. I provided reasons for thinking that it is extrinsic and showed how on this hypothesis Fodor's argument is not sound. I also showed how Fodor's own defense of the causal efficacy of some extrinsic properties (such as being a planet) fails.

CHAPTER FOUR: AN ACCOUNT OF EXTRINSIC CAUSAL POWERS

4.1 INTRODUCTION

In this chapter I turn to the second task of my dissertation: to outline an account of how wide semantic properties can be robust causal powers. Chapters 2 and 3 comprised the first task: to show that none of the a priori arguments in the relevant literature demonstrate that wide contents can't be robust causal powers. Overall, my strategy is to show that whether they are or not is entirely an empirical matter. I argued in Chapter 2 that the arguments presented there relied implicitly on the Non-Locality arguments, and in Chapter 3, I found that the Non-Locality arguments did not succeed. I'll review the main arguments of Chapter 3 before proceeding.

The Non-Locality arguments, presented mainly by Fodor and McGinn, are based on an intuitive conflict between the spatiotemporal locality of causal interactions and the extrinsicness of wide contents. It is argued that extrinsicness entails two kinds of property non-locality that are incompatible with the locality of causal interactions. Contents are either not instantiated "locally" by particulars, or a token's content properties can differ when its "local" properties do not. In the first case (McGinn's), contents are simply not in the right place to be properties in virtue of which particulars enter causal interactions. In the second (Fodor's), a content-based taxonomy is not a taxonomy by genuine causal powers because wide contents do not strongly supervene on the agent. Real causal powers can differ only if "local microstructure" does, but widely individuated mental states can differ when "local microstructure" does not.

I argued in response that contents do not violate either type of property non-locality. I first motivated my responses by showing how the available answers to Fodor

and McGinn do not preserve a robust, rather than merely proleptic, causal role for semantic properties. It may be granted that a property must play a causal-explanatory role in a successful research program to be a causal power, and it may be granted that semantic properties play such a role in an ongoing scientific psychology. But, for Fodor and McGinn, this is not sufficient for a property to be one in virtue of which a token participates in a singular causal interaction. Satisfying both kinds of property locality are also necessary. Burge in particular disagrees with this claim. I argued, *pace* Burge, that these additional criteria must be met to avoid the relegation of intentional causal powers to the second-class status of merely causally-explanatory properties – properties that at best point to other properties that do the real causal work. McGinn's and Dretske's own responses to the Non-Locality arguments, and Fodor's (1994) defense of contingent covariation between syntax and semantics, bear out this assessment.

I then argued that neither Fodor nor McGinn succeeds in showing that extrinsic contents violate either type of property locality. *Pace* McGinn, externalism about content does not entail that contents are not instantiated locally. McGinn appears to confuse what is constitutive of a content type (e.g. what it is to be an extrinsic water-belief) with what constitutes the instances of a given content type (or of its supervenience base). It is compatible with externalism that content properties are fully constituted in a particular's microphysical properties, and so are "local" in McGinn's sense. What the externalist must deny is that the *identity* of these instances is determined by the aggregate of the microphysical properties. The "extended mind" thesis (or vehicle externalism), which claims that at least some of the constituents of mental types are located outside the head, is compatible with content externalism but is not entailed by it, as McGinn seems to hold.

Pace Fodor, I argued that the validity of his argument depends on the sort of hypothesis one accepts regarding the nature of local microstructure. According to the intrinsic hypothesis, which Fodor holds, local microstructure is the intrinsic component of wide content (or of its supervenience base). I also argued that Twin Earth shows (and only shows) that wide contents differ with differences in molecular structure. Given the intrinsic hypothesis, then we can validly conclude that wide contents will not differ when causal powers do.

The extrinsic hypothesis holds that local microstructure is extrinsic. On this view, even if it is true that causal powers differ when local microstructure differs, local microstructure differs when wide content does. If so, then Fodor's argument is invalid. The fact that wide contents can differ when molecular structure differs may be true, but it doesn't give us any reason to draw Fodor's conclusion.

I also responded to a likely objection. If I am right, we would be classifying Oscar's and Twin-Oscar's brains as having different causal powers. But surely, the objection would go, molecularly identical neural structures will behave the same; thus, our causal laws ought to subsume them equally. I argued that this objection fails because even molecular identity does not guarantee identical behavior. Chemistry taxonomizes mirror-image (left- and right-handed) molecules as distinct because their causal powers can differ significantly in kind as well as degree: the left-handed enantiomer of ketamine (Ketalar) is an anesthetic, the right-handed form is an excitant that can cause psychic disorders.¹⁹³ So it is neither descriptively nor prescriptively true of science that it

¹⁹³ Jacques (1993: 106-7) notes that in cases where enantiomers of a given compound have qualitatively different properties, "it is neither inconceivable nor uncommon for one and the same molecule to serve multiple purposes: as an antirheumatic, a pain-killer, an antipyretic, and so on. If it can be resolved [separated into right- and left-handed forms], one of the enantiomers might be analgesic and the other more

exclusively classifies or ought to classify particulars under intrinsic properties for causal-explanatory purposes. Science seeks generality, but equally tries to avoid overgenerality.

My plan in this chapter is to take a closer look at causal powers in science and then to apply these results to the case of wide content. I'll show that science is open in general to robust extrinsic causal powers, and then show how this solves (more precisely, dissolves) the problem of the causal efficacy of wide content. The chapter has two main parts. In the first part, I use the case of molecular structure to show that the philosophical prejudice against robust extrinsic causal powers has no scientific basis, and also suggest an explanation for the prejudice. In the second part, I argue that in the computational theory of mind the class of properties now lumped under the term "syntax" (in the lower-level sense) may include both intrinsic and extrinsic properties, and some of the latter may be identical to or be the supervenience bases of wide semantic properties. Since "syntax" is generally agreed to be robustly causal, I conclude that the problem of externalism is an artifact of the assumption that all structural (a.k.a. "syntactic" or "local microstructure") properties are intrinsic.

I then show how the possibility of semantic structural properties fits into an account of robust mental causal powers that has been defended by Derk Pereboom. Pereboom argues that some mental properties may be irreducible by analogy with the multiply realizable structural properties of complex artifacts. While he assumes such

specifically antirheumatic. This type of preparation is unquestionably one in which resolution is a must, if only because patient sensitivity to these different effects ("side effects") are largely unpredictable." Other interesting chiral examples (also from Jacques): only the left-handed form of methyl dopa (Aldomet) is efficacious for treating high blood pressure; and the two enantiomers in propranolol (Avlocardyl), a beta-blocker, also have quite different properties. Most drugs today are in racemic (mixed) form, but efforts to synthesize "pure" forms which include just the active form are underway. Jacques notes that one researcher has recommended that a warning – "Contains 50% isomeric ballast" – should be imprinted on drug packages, "since the products of the transformation of the two components of a racemate in the organism may be different and not have the same long term effects."

properties are intrinsic, the assumption is inessential to his model. So his account has room for the possibility that some irreducibly mental structural properties are both extrinsic and semantic.

4.2. EXTRINSIC CAUSAL POWERS IN THE PHYSICAL SCIENCES

4.2.1. The Extrinsicness of Molecular Structure

Some theoretical chemists argue that the concept of a structured molecule can only be explained consistently with quantum mechanics if structure is an extrinsic property, one which essentially involves the molecule's interaction with its environment. (This claim will be made more precise in what follows.) Notably, there is no accompanying suggestion that the extrinsicness of molecular structure would alter the causal and taxonomic role it plays in chemistry. From chemistry's point of view, the shocking aspect of the argument is the implication that chemists are deeply mistaken about the nature of molecules. This is particularly disturbing because the interpretation of experiments with molecules in terms of classical, intrinsic structures is extremely reliable; it is, in the words of the critic discussed below, "the central dogma of molecular science."¹⁹⁴

The debate over molecular structure arises within the context of the problematic relation between classical chemistry and quantum mechanics.¹⁹⁵ It is the most dramatic of a number of revisions of the foundational concepts of chemistry proposed or made since

¹⁹⁴ Woolley (1978: 1074). I discuss Woolley in detail below.

¹⁹⁵ Claverie and Diner (1980: 54), whose assessment of Woolley will be discussed below, make this point explicitly: "The classical concept of molecular structure (namely, a set of atoms with a well-defined geometrical arrangement in space) originated well before the advent of quantum theory, and, contrary to widespread views, these two conceptual schemes (classical molecular structure and quantum theory) are not easy to reconcile. ... [W]e deduce that the "classical structure" concept is indeed non-trivial from a purely quantum-mechanical point of view; ... it appears that, at the present time, none of these classical concepts and behaviors is convincingly derivable from a strict quantum mechanical basis."

the advent of quantum mechanics. It should be emphasized that the debate over molecular structure is far from over. But the existence of the controversy and the radically different views of molecules that are emerging from it suffice to raise serious doubts about the conclusions of philosophical arguments involving intuitions regarding molecules. I will argue in the next section that the claim that wide semantic properties are not causal powers is one such conclusion: it is an artifact of intuitions about local microstructure that are rooted in analogies to classical molecular structure. In this section, however, I will present one of the seminal arguments against the classical model of molecular structure and consider some philosophical objections and implications of the debate.

What is the Classical Model? The classical molecular model – “classical” in the sense of classical or Newtonian mechanics – is deeply embedded in the chemical sciences as well as in popular intuition.¹⁹⁶ We think of a molecule as a collection of distinguishable atoms whose nuclei have different fixed positions in the molecule and which interact with each other mainly through their electrons. More precisely:

[T]he molecule is considered to be a quasi-rigid assembly of atoms, connected by bonds, and the body thus formed, which has a definite size and shape, performs continuous (classical) rotational and translational motion in the course of time. In a somewhat more refined picture, more closely related to a quantum description, one speaks of a quasi-rigid framework of nuclei embedded in an electronic cloud, that takes into account the vibrations of this nuclear framework around its equilibrium configuration. The essential element of this picture ... is the idea of a classical-type rotational and translational motion of this nuclear framework in the course of time. ... [This is] the idea that, at any instant of time, the geometrical structure of the nuclear framework *actually exists* with some definite orientation and position, even when we do not perform any measurement. Accordingly, this supposedly actual instantaneous structure would move classically in the course of time as a familiar rigid body. (Claverie and Diner 1980: 60-61)

¹⁹⁶ Ramsey (1997: 234) uses the phrase “folk molecular theory” to describe the intuitive view of molecules based on classical chemistry. As will become clear, these intuitions really are classical. Claverie and Diner (1980: 54) also make this point: “Despite its microscopic nature, the molecule is a classical concept as far as the problem of its spatial structure is concerned, and could be called an example of classical behavior on a microscopic scale.”

If this model is accurate, then the term “structure” denotes a nuclear framework – a sort of microscopic inner scaffolding. Structure exists as a kind of object – a rigid body whose frame is composed of nuclei – and it exists independently of our observing it.¹⁹⁷

This model is classical in the strict sense of classical mechanics – not merely because it is the “received” view or is deeply embedded in intuition – because it treats the nuclei in molecules as classical particles. (The root of Woolley’s and others’ criticisms is the demand that nuclei be treated fully quantum mechanically.) This is because they are assumed to be fixed in position within the molecule (allowing for a bit of vibration). For example, a centrally located carbon nucleus stays in the center and a peripheral carbon nucleus remains peripheral even as the molecule rotates and translates through space, and even if the molecule is isolated – that is, not interacting with the environment in any way.¹⁹⁸ The fixed positions of nuclei make their behavior classical in the following ways.

First, it implies that in principle nuclei still have trajectories, or paths through space at any point of which a particle has both a definite position and a definite momentum. This violates the fact that quantum particles (unlike classical particles) cannot have a definite position and momentum simultaneously.¹⁹⁹ Second, nuclei in fixed

¹⁹⁷ A rigid body is a body or system of particles in which the distances between the particles do not vary. Rigid body motion is motion that occurs when forces applied to a rigid body do not deform the body in any way. The rigid body assumption is literally false and is known to be false for all composite bodies; in quantum mechanics, it is even false at absolute zero. But it expands the scope of classical laws beyond just explaining the motions of two atoms (although at that level Newton’s laws are false anyway).

¹⁹⁸ An isolated system is one that is not interacting – not exchanging energy or mass – with its environment. A closed system is one exchanges energy but not mass, and an open system is one that can exchange both.

¹⁹⁹ See Patsakos (1967) for discussion of classical vs. quantum particles, and Liboff (2003). In quantum mechanics, some observables, including position and momentum, and time and energy, are complementary: the more definitely we can specify the one, the less definitely we can specify the other. More precisely, complementary variables do not commute: if A and B are complementary, then $A \times B$ will not be equal to $B \times A$. The Heisenberg uncertainty principle involves our expectation of the value we will observe in a particular measurement of a complementary variable. When we say one of a pair of complementary variables is certain at a time, we are saying that we know precisely what value a measurement of that observable at that time will yield. In this case the other complementary variable is maximally uncertain: an

positions are by hypothesis immune to the usual consequences of wave-particle duality.²⁰⁰ Quantum, but not classical, particles have wave characteristics and as such are subject to “delocalization” (or “spreading”): over time we are less able to determine a precise position of a quantum particle and more likely to find it somewhere in an ever-widening range of positions. There is no such thing as “delocalization” in classical mechanics. So if nuclei in a molecule are highly localized – that is, if they have and maintain sharply peaked probability distributions for measurements of their position over time – this would be a matter of empirical verification. It would not generally be the case, as it is by assumption in the classical model.

Third, and most importantly for Woolley’s argument, identical particles in quantum, but not classical, mechanics – such as two carbon nuclei or two electrons – are fully indistinguishable and interchangeable. In particular, they are permutationally symmetric: when we make a measurement of position of nuclei within a molecule, it cannot matter which of two identical nuclei in a given system we observe. Identical classical particles, in contrast, can have distinct physical descriptions if they occupy distinct determinate positions in the system relative to each other. For example, in a rotating molecule a centrally located carbon nucleus will have smaller angular

immediate subsequent measurement of that observable may yield *any* of the possible values it may take with equal probability. Time and energy are also complementary: if we know precisely what value for the energy we will observe, it is completely unknown how long it will take to observe it. Conversely, the longer the time-frame of observation, the less our expected value for its energy will deviate from the measured value. At infinity, our expected value for the energy is right on the money.

²⁰⁰ See Atkins (1998: 294) and Liboff (2003: 54). Atkins remarks: “Duality strikes at the heart of classical physics, where particles and waves are treated as entirely separate entities. ... Quantum mechanics acknowledges the wave-particle duality of matter by supposing that, rather than travelling along a definite path or trajectory, a particle is distributed through space like a wave.” Liboff adds: “When an electron (or photon) exists within a well-defined locality of space (momentum is ill-defined), it acts very much like a particle. ... When the electron does not exist in a well-defined locality of space, its momentum can be defined more precisely. Under such circumstances the wave character of the electron manifests itself. We cannot follow *it*. A whole wave is propagating.” The trajectory of a system in classical mechanics has been replaced in quantum theory with the wavefunction ψ , which contains all the dynamical information about a system. Identical particles have identical wavefunctions.

momentum, hence a distinct physical description, than a peripherally located carbon nucleus.²⁰¹ We can also follow the movement of each distinct carbon nucleus through space and time. This cannot generally be done in quantum mechanics.

The classical treatment of nuclei in this model contrasts sharply with that of the electrons, which are treated fully quantum mechanically. For example, nowadays electrons no longer have orbits, which are trajectories. They “occupy” orbitals, which are basically probability distributions. To say that an electron is in a 1s orbital is to say that there is a 90 percent probability that in a given measurement an electron will be observed within a sphere-shaped region relatively close to a nucleus. Electrons are also delocalized and interchangeable; if an electron “occupies” the 1s orbital, it makes absolutely no difference *which* electron we observe as long as we find *an* electron somewhere in that region when we make a measurement. These differences are reflected in the metaphorical description above of electronic motion in terms of a “cloud”.

It should be clear that the classical explanation of molecular structure, quite apart from issues about its nature, is under pressure due to the quantum mechanical treatment of electrons and electronic motion. This is because molecular structure is explained in terms of nuclei that are linked by bonds, while bonds are explained in terms of interacting electrons. They are (classically) quasi-rigid linkages formed by electromagnetic forces acting between electrons orbiting in the outermost (valence, or highest-energy-level) shells of atoms.²⁰² However, if electrons “occupy” orbitals, at best bonds can only be

²⁰¹ As Woolley (1978: 1077) explicitly notes, “it is characteristic of the [classical] molecular structure description that identical nuclei can be dynamically inequivalent.”

²⁰² More precisely, classical structure is explained in terms of the arrangement of electrons in a molecule that minimizes repulsions between the pairs of electrons that form the bonds. Differences in molecular structure depend on differences in bond lengths, strengths and angles. Bond-forming electrons are in the valence or outermost shells of atoms – they occupy the highest energy levels. In valence theory, the valence electrons of a central atom minimize repulsion by taking positions on opposite sites of the atom; in

probability maxima where orbitals overlap. Strictly speaking, we can't say that a BeF₂ molecule *is* linear; we can only say that there's a 90 percent probability that we'll observe a linear arrangement at a given instant if we make a measurement of it.

As we shall see, if critics of the classical model are correct, we cannot even say that much. Instead, we can only say that there's a 90 percent probability that we'll observe an emission or absorption spectrum that is *consistent with* the existence of a molecule with a classical structure of the linear type. It doesn't follow that it actually has this structure in the way the classical model posits.

Woolley's Argument Against Classical Structure. The *locus classicus* of the debate about molecular structure is an argument by R.G. Woolley (1978) that has been developed or augmented by him and others in the years since.²⁰³ It must be emphasized that Woolley takes aim at the nature of molecular structure, not its taxonomic or causal-explanatory role.²⁰⁴ His basic argument may be stated non-technically as follows:

- 1) According to the classical model, nuclei in a molecule have fixed relative positions in it, whether or not it is isolated.
- 2) Identical particles with distinct fixed positions in a system – such as identical nuclei in the classical model – can have distinct physical descriptions.
- 3) In quantum mechanics, identical particles in a system cannot have distinct physical descriptions.
- 4) Observed molecules can be described classically – that is, in terms of a fixed nuclear framework in which identical nuclei can have distinct physical descriptions.
- 5) Our observations are consistent with QM if molecules have classical descriptions only when they being observed or are otherwise not isolated.
- 6) It follows that isolated molecules cannot be classically described, and the classical model is false.

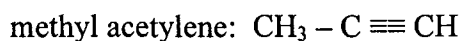
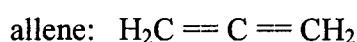
consequence the other atoms will be arranged linearly (as in BeF₂). The concept of a bond and valence theory are also being questioned in the chemical literature; e.g. Weininger (1984) notes that there appears to be no quantum mechanical justification for the classical assumption that electrons bond in pairs.

²⁰³ Other arguments attacking classical molecular structure include Primas (1980) and Ramsey (1997, 2000), summarized below. Claverie and Diner review the debate and end up siding with Woolley.

²⁰⁴ Woolley makes this point explicitly (1985: 1082): "The importance and usefulness of the molecular structure concept is not an issue here; no one is suggesting that it be abandoned. What is of concern is our *understanding* of the structure concept in the light of current knowledge of physics."

Strictly speaking, the argument only points out an inconsistency between the classical model and quantum mechanics. However, to at least the extent that quantum mechanics supercedes classical mechanics, if Woolley is right then there is only one way to jump.²⁰⁵ Moreover, to say that the classical model is false is not yet to say what is true. The ongoing debate still veers between anti-realism and realism, and among the realists about structure there is no one dominant view.²⁰⁶ They do however share the claim that molecular structure, if real, is extrinsic: a molecule's having a true classical description entails that it is not in an otherwise empty world. (I will discuss the entailment below.) To make this claim more concrete, I will sketch a realist yet extrinsic view of molecular structure which is consistent with remarks made by Woolley and some other chemists. But I cannot claim that this is Woolley's considered final position.

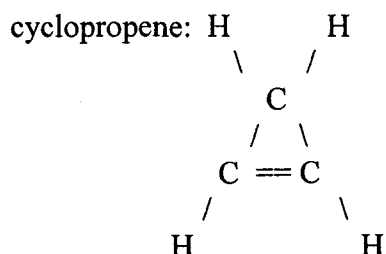
The thrust of Woolley's argument may be illustrated most clearly in the case of isomers, or chemical species that have the same atomic formulas yet different classical structures. The isomers of C_3H_4 , for example, are sometimes represented as follows:²⁰⁷



²⁰⁵ Weininger (1984), who sympathetically examines Woolley's argument, takes care to note that there is always the possibility of choosing to jettison quantum mechanics.

²⁰⁶ For example, Primas (1980) defends a radically anti-realist position in which the system/environment distinction is arbitrary; each partition of holistic nature "creates its own reality". (See Weininger (1984: 942). Claverie and Diner (1980) defend a non-classical realist position. Ramsey (1997) formulates a "perspectival realist" view in which structure is dependent on the time-frame of observation (or interaction). Woolley has been equivocal, but has cited Claverie and Diner's work with approval.

²⁰⁷ The C_3H_4 example (but not the use to which I put it) is from Ramsey (1997: 237).



These isomers are taxonomized as distinct chemical species because they have distinct classical structures and play distinct causal roles: that is, chemistry taxomizes its kinds in terms of structural differences, which explain differences in chemical behavior. However, if Woolley is correct, isolated C_3H_4 molecules have identical physical descriptions; they cannot be said to belong to any determinate chemical species.²⁰⁸ A C_3H_4 molecule will only have a classical structural description – associated with either allene, or methyl acetylene, or cyclopropene – and thus be taxonomized for causal-explanatory purposes, when it is interacting with the environment in some as-yet unknown way. In sum, a molecule would not be a microscopic rigid body; it would be an interacting cloud of nuclei and electrons which in the right environment has a determinate chemical character.

It should be emphasized that the role of the environment on this alternative model is not that of causing a classical structure to exist. This would misread Woolley's criticism. On his view, a classical-type description either applies or does not apply to a fully quantum system. His claim is that the classical-type description will be true of the quantum system only when it is not isolated. We would continue to use these descriptions

²⁰⁸ Woolley (1985: 1083) emphasizes this point: "[T]he full spin-free, nonrelativistic molecular Hamiltonian describes the interactions of an assemblage of electrons and nuclei. *It does not describe a particular molecular species* [his italics]. The Hamiltonian shared by, for example, cubane and cyclooctatetraene is also shared by vinylbenzene or any other system where eight carbon nuclei, eight protons and 56 electrons are interacting. Its stationary state wavefunctions describe the correlations between the electrons and nuclei due to their Coulombic interactions, but they do not lead to anything that can be called classical molecular structure."

(and the term “structure”) just as we now use the term “particle” despite the differences between classical and quantum mechanics. It would not follow that any of these terms refers to a classical entity. It would mean only that molecules, like particles, sometimes exist in ways that can be described classically.

A plausible realist yet extrinsic model of structure might hold that our classical-type descriptions, if they apply at all, describe a process, not a rigid framework, which is an object. As noted, bonds are, ultimately, electromagnetic forces interacting among a molecule’s constituent particles. The extrinsic alternative adds interactions with the environment to these interactions among nuclei and electrons. On this view, classically-describable structure is a dynamical state of an embedded molecule. This is consistent with the characterization in the literature of molecular structure as a “response” property or a secondary property, and with arguments in which structure is relative to the time-frame of observation (or interaction).²⁰⁹ The taxonomic importance of the external component of this complex process lies in the fact that it accounts for the difference between (e.g.) allene, methyl acetylene and cyclopropene, given that these distinct chemical species have identical atomic constituents. So if a molecule has a particular classically describable structure, it is not in an otherwise empty world. In the future we may be more specific about the nature of that universe for any given type of chemical species. But this is already sufficient for the claim that molecular structure is extrinsic.

A word should be said about the relation between Woolley’s conclusion and chemical experiments. Woolley agrees with his opponents that the nuclei in molecules observed in most laboratory setups do largely exhibit behavior that for decades has been

²⁰⁹ Claverie and Diner (1984) use the term “response property”, while Ramsey (1997) uses the term secondary property. Ramsey (2000) emphasizes the role of time in characterizing structure.

reliably interpreted in terms of the classical model. But the experimental evidence for classical molecular structures actually consists of inductive inferences from indirect measurements of internal molecular energy levels. In the various forms of spectroscopy, for example, molecules are bombarded with radiation in a range of frequencies. In this interaction, the particles in the targeted molecules undergo transitions to new energy levels, absorbing and emitting photons (quanta or particle-like packets of electromagnetic energy) of different frequencies or energy levels. These varying photon frequencies are recorded on a spectrum, which constitutes an indirect record of the internal dynamics of the molecule. Correspondences have been established by induction between different spectral frequencies and different classical bond characteristics.

What Woolley points out is that one can accept that the spectral lines do accurately record the dynamical state of the observed molecule without also making the inference to a classical structure.²¹⁰ And on his view we are not motivated to make this inference if it implies a contradiction with quantum mechanics. Instead, we need only reconcile what quantum mechanics tells us about the behavior of identical particles with the spectra of observed molecules that justify classical descriptions. The obvious, and explicitly quantitatively tractable, solution is to say that the process of observation itself – or some type of environmental interaction more generally – is responsible for the behavioral distinctions within molecules that give rise to classical-type descriptions.²¹¹

²¹⁰ Woolley (1985: 1083) remarks: “[I]t is an elementary point of logic that seems easily overlooked that the success of this [classical] model in accounting for chemical observations *does not imply* that the molecular structure hypothesis is a *true* account of the nature of chemical substances as they “really” are (of “reality”); to go down that road is to become entangled in the mirages of metaphysics. The metaphysics of classical realism is optional.”

²¹¹ For example, Claverie and Diner (1980: 64) argue for extrinsicness but against measurement-dependence, as does Ramsey (2000: 119). Woolley and others also suggest that since we observe the appearance of classical-type structures over time, *continual* environmental interaction is required for molecular structure. Claverie and Diner (op.cit.: 68) explicitly consider Woolley’s suggestion a

That is, our *total* evidence, including both quantum mechanical calculations and observed spectra, is more adequately explained by this hypothesis than by the classical model.

In response to Woolley's argument, it might be objected that the inconsistency between the classical molecular model and quantum mechanics only arises when we consider what they say about molecules *in isolation*. Molecules either can or cannot have classically describable structures in isolation. But molecules do not exist in isolation in nature. If all we ever observe are classically describable molecules, who cares? At worst we have two alternative hypotheses which agree on the observables and disagree for cases that we cannot observe.

As an initial response to this objection, I note that the fact that there are two alternative hypothesis to explain our observations is sufficient to undermine any *a priori* claims about the intrinsicness of molecular structure. Moreover, the extrinsic model agrees with quantum mechanics, whereas the classical model does not. This lends more support to the extrinsic hypothesis, given that a better hypothesis is one that is consistent with existing well-confirmed theories. However, there are additional physical and philosophical reasons why the argument is not trivial even though molecules, like physical systems generally, do not naturally exist in isolation. These reasons support a realist yet extrinsic view of molecular structure.

First, the assumption of isolation is made by both sides in this debate. It is an assumption that is widely employed in physics to facilitate physical calculations, such as of a system's motion. It eliminates having to take into account the effects of external

quantitatively tractable answer to the problem of explaining molecular structure, and divide that problem into two: explaining the appearance of classically describable structures, and explaining the maintainance of these classical-type structures (with localized nuclei) over time. Their answer to both questions is environmental interaction.

interactions between the system of interest and other systems. This is why efforts to study isolated molecules experimentally preceded the debate about molecular structure and have continued alongside it.

The assumption of isolation is also independent of any assumptions one might make about a composite system's internal dynamics. One is free to assume not just that a molecule is isolated but that the nuclei in the isolated molecule are relatively fixed in position. This in fact that is how chemists usually proceed. They simplify their calculations of molecular energy by adopting some type of rigid-body assumption insofar as nuclear motion is concerned.²¹²

Now natural composite systems are no more rigid than they are isolated. But the rigid-body assumption, unlike the isolation assumption, is *not* neutral in this debate, since it is an assumption about the motions of constituent particles. In particular, it sustains the classical molecular model's rigid nuclear framework. When we do not make the rigid-body assumption at the microscopic level, in effect we no longer ignore the quantum nature of nuclei. This is precisely what Woolley does: he continues to assume isolation in order to obtain as exact a calculation as possible of a system's total energy; but he does not *also* assume that the nuclei are fixed in position. Since identical quantum particles in a system are fully interchangeable, the result is that there is no justification from basic quantum principles for a classical-type description. The conflict with the classical model does not arise because Woolley assumes the molecule is isolated, but because he does not assume that the nuclei form a rigid body.

²¹² The most common assumption is the Born-Oppenheimer approximation, in which nuclei are assumed to be fixed at given relative distances and the electronic motion is calculated. Repeated calculations done with different estimations of nuclear distances result in a curve. A minimum on the curve represents an equilibrium bond length. See Atkins (1998).

Second, since our experimental setups increasingly can achieve virtual isolation, Woolley's argument is hardly academic.²¹³ This means we will soon have (if we do not already) setups that will allow us to test the distinct observational consequences of the classical and non-classical models for isolated molecules. (In fact, some chemists argue that molecular structure is not intrinsic because we observe different structures of the same molecules under different conditions of observation.²¹⁴) The classical model tells us that achieving isolation will give us better resolution of classical structures. If Woolley is correct, the classically describable structures will disappear.²¹⁵

Clearly, the direction of future molecular research would differ if non-isolated molecules do not have classical structures, even if they always satisfy classical-type descriptions. For example, theoretical chemists following Woolley have suggested models of molecules which aim to distinguish and quantify the types of interactions that may be required for a given type of classically describable structure.²¹⁶ This research would not even make sense if structure is classical. This is why the question Woolley

²¹³ Woolley (1985: 1083) has likened his use of the isolation assumption as being on a par with the role of the principle of inertia in mechanics, as when Galileo assumed a body that moves without the influence of external forces in order to develop his new science of dynamics. Conditions of "microgravity" or free fall are a subject of current NASA research, with experiments involving falling objects in satellites in orbit.

²¹⁴ See Ramsey (2000) for an argument along these lines.

²¹⁵ As Weininger (1984: 940) remarks: "Conventional wisdom tells us that by isolating a molecule and more finely resolving its energetic transitions we should be able to increase our detailed knowledge of its structure. However, if the analysis of Woolley and others is correct, then we may learn less about a molecule's structure by isolating it than we do under less extreme experimental conditions."

²¹⁶ See Claverie and Diner (1984), who suggest that weak interaction with a vacuum electromagnetic field or with neighboring molecules may be necessary for classical-type structures. They also suggest that in isolation molecules may be said to have what they call a "quantum structure": "[I]f, at any instant of time, we should measure the positions of the nuclei of some molecules, we would find some (relative) configurations of the resulting positions with higher probability, and those (relative) configurations with maximal probability clearly enable us to speak of some underlying structure." They add that this quantum structure "is, admittedly, of a less pictorial nature than the 'naive' structure of the classical molecular model", but it might be "the 'germ' of an account of the emergence of a classical-type structure in suitable environmental conditions."

ultimately raises is whether the classical model is still scientifically fruitful rather than increasingly irrelevant to explaining chemical behavior.²¹⁷

Third, conditions of isolation are physical counterparts of the conditions which determine the difference between intrinsic and extrinsic properties. For simplicity, we think of that difference in Lewis's terms: an extrinsic property is one that implies accompaniment, whereas an intrinsic property is compatible with either loneliness or accompaniment.²¹⁸ This means we must consider what is the case in a lonely world to determine whether a property of an object is intrinsic or extrinsic. An isolated molecule is a lonely molecule. Although this is an unnatural physical possibility, it is the exactly the kind of situation which we would expect to have to consider to determine whether molecular structure is intrinsic or extrinsic.

Of course, we can also ask about metaphysical possibility: could there be, *metaphysically*, an otherwise empty world in which molecules have classically describable structures? In other words, is it really correct to say that a molecule's having a given classically describable structure *entails* that it exists in a non-empty universe? Is it really a matter of the *nature* of molecular structure that it is extrinsic? I think there are two strong responses to this doubt.

First, the doubt depends heavily on our intuitions about what might be the case. But our intuitions about molecular structure are deeply influenced (or simply infected) by the classical model. It seems as clear as day to us, for example, that water is (identical to)

²¹⁷ As Woolley (1978: 1078) states: "Naturally I recognize that since much of the quantum theory appropriate here has yet to be worked out, experimentalists will regard this critique of molecular theory as being of little direct help to them; equally one hopes that these experiments are still performed to investigate primarily the properties of matter, which need not be the same as "the determination of very precise molecular structures", and without some agreement about the theoretical principles that will be useful in the last quarter of the 20th century, this area of physical chemistry may well find itself up a blind alley as its traditional molecular models become less and less relevant to contemporary experiments."

²¹⁸ See Lewis and Langton (1998) and Francescotti (1999).

H₂O. Most of us also accept that this is a necessary *a posteriori* truth: we discovered the nature of water after it had already been picked out by its behavior (and other features). Chemists determined what the basic elements are, and now we know that water *could not* but be H₂O.

But chemists also tell us that water has a polar V-shaped structure that explains much of its chemical behavior. This configuration is as much a part of the nature of water as its atomic composition, given the fact that chemistry taxonomizes its kinds by structure and not by mere atomic composition. So if we are more precise about our *a posteriori* necessary truths, we should say that water is (identical to) two hydrogen atoms and one oxygen atom whose particles are interacting in a particular way. Similarly, we would say that allene is (identical to) three carbon atoms and four hydrogen atoms whose particles are interacting in a particular way (a way that is different from that of methyl acetylene and cyclopropene). “Interacting in” is more precise than just “configured in” because a configuration is a matter of bonds, and bonds are electromagnetic forces acting between molecular constituents.

Now if Woolley is correct, chemistry is now telling us that the characteristic structural descriptions of molecules physically cannot apply to molecules in isolation. This is the point at which the questions about the nature of structure arise. Even if Woolley is correct physically, *could* there be a C₃H₄ molecule in an otherwise empty universe to which the classical description appropriate to allene applies?

We can surely imagine this to be the case. But these intuitions have got to be suspect, because they are *classical*. They express precisely what the classical model claims (see premise 1 of Woolley’s argument). Therefore they cannot be considered

neutral evidence in this debate. If Woolley is correct, and if we knew more about the natures of the environments in which molecules must exist in order to have true classical-type structural descriptions, then it is quite plausible that these classical intuitions would be replaced by others. We would then intuit that a molecule that truly satisfies the classical structural description characteristic of allene could not exist in an otherwise empty universe. Our updated and corrected necessary *a posteriori* truths would be of this general form: water is (identical to) two hydrogen atoms and one oxygen atom interacting in a particular way with each other and with an environment that has a particular nature.

It is worth noting in this regard that chemistry was initially a science of proportions, not structured molecules.²¹⁹ The classical model was adopted over initial resistance to a large extent in order to account for the distinct causal powers of isomers. After chemists discovered around 1830 that silver cyanate (AgCNO) decomposes when heated while silver fulminate (also AgCNO) explodes violently when heated, it became necessary to draw taxonomic distinctions that would reflect these differences in behavior. Classification by composition alone was not fine-grained enough for causal explanation. The resulting structure-based causal-explanatory taxonomy, developed before quantum mechanics and spectroscopy, still prevails. From this perspective, Woolley's argument shows that chemistry may need an even more fine-grained causal taxonomy than just classification by configuration, beyond or including the adjustments already being made in the light of chirality. But before 1830, questions about the metaphysical nature of molecular structure could not have arisen, and no one would have had any intuitions about it at all. If, as realists, we dismiss this pre-1830's lack of intuitions as stemming

²¹⁹ See Weininger (1984: 941), and Le Poidevin (2000) for a more extensive discussion.

from ignorance about the nature of molecules, we should be prepared to accept that today's classical intuitions also stem from ignorance.

A second reply to the problem of whether molecular structure is extrinsic by nature is to claim that chemical kinds are natural kinds, and that the nature of natural kinds is a matter of physical, or perhaps nomological, necessity. On this view, we simply do not consult our modal intuitions at all on the basis that modal talk is meaningless. So what might be the case metaphysically is irrelevant. On this view, Woolley's claim is that molecules physically can't have classical-type structures if quantum mechanics is correct, and we needn't make any further claims about the nature of structure. We would still say that molecular structure is extrinsic, but we would use a local notion of extrinsicness. (A property is locally extrinsic iff it is an extrinsic property of some individual x that has it; a property is globally extrinsic iff it is extrinsic for all individuals that have it.)

Philosophical implications. There are two implications of the debate about molecular structure that are most pertinent in this context.

First, the debate shows that any philosophical argument in which *a priori* intuitions about molecules play a role must be reconsidered. Overwhelmingly if not exclusively, these intuitions are strictly classical. We already saw that due to chirality we cannot simply assume that because Oscar and Twin-Oscar are molecularly identical, they therefore behave (more precisely, move) the same. We must also know that they are or are not mirror-images, and being a mirror-image (e.g. either being right- or being left-handed) is extrinsic. Similarly, if molecular structure is extrinsic, then even if Fodor were right (in his Non-Locality argument) that local microstructure supervenes on molecular structure, it would not follow that both are intrinsic. Fodor's argument would not give us

any reason to think that higher-level extrinsic properties (such as wide content or, if the extrinsic hypothesis is true, local microstructure) are not causal powers.

Second, some quantum chemists assessing Woolley's proposal about molecular structure explicitly consider it both quantitatively tractable and scientifically fruitful. The appeal to external determinants of structure, and hence the recognition of chemical kinds as extrinsic, is made *because* the structure concept is indispensable for explaining differences in molecular behavior. This shows that extrinsic properties do not inherently violate scientific demands for explanatory generality or fecundity and are not inherently unsuitable for playing a role in causal law, even in the quantitative physical sciences.

It follows that whether an extrinsic property is a causal power or not can only be determined by empirical research, not *a priori* considerations. Coupled with my arguments of the previous chapters, this means that there are no *a priori* reasons to think that the causal powers of wide semantic properties aren't as robust or genuine as they seem to be. If quantum chemists have no qualms about robustly causal-explanatory extrinsic properties, scientifically-minded psychologists should feel free to assume, as they conduct their research, that the *prima facie* causal role of wide contents is genuine.

4.2.2. The Resistance To Extrinsic Causal Powers

As noted, nowhere in Woolley's discussion, or that of others assessing his work positively, is it suggested that because molecular structure may be extrinsic, it can no longer to play its central causal-explanatory role.²²⁰ It is far more important to save the molecular structure concept, taking its causal-explanatory viability for granted, than to

²²⁰ Ramsey (2000) suggests that while a molecule may have some structure ("shape") or other that is causally explanatory in every situation, the structure may differ in different explanations. This would not deprive structure of its role; it would only make our explanations more limited in scope.

save the assumption that it is intrinsic. So we might wonder where the resistance to extrinsic causal powers, at least among many philosophers if not also many scientists, comes from.

For philosophers, there may be historical reasons tracing back to Locke's distinction between primary and secondary qualities and his reservation of causal powers to the primary qualities.²²¹ I'd like to suggest a non-historical reason that stems from the nature of science itself. The basic idea is that it is standard scientific practice to simplify, idealize away from or ignore aspects of the physical world in order to expand the scope of our laws and make them easier to use. As an unintended consequence, this practice may mislead us about the nature of some of the properties actual phenomena have. In particular, it may seem as if causal-explanatory properties are always intrinsic, when this appearance is just an artifact of ordinary scientific practice.

One example of how we simplify, or idealize away from, actual phenomena can be found in Cartwright's influential work on physical law. Cartwright (1983: 54) argues that fundamental laws of physics are literally false as descriptions of facts, but lose their wide explanatory scope when amended by *ceteris paribus* clauses to be true. For example, the classical law of gravitation (as usually stated) subsumes any two massive bodies. But many bodies also have electric charge, and electricity also exerts forces between bodies, as described in Coulomb's law. The resultant force between two massive and charged bodies is not explained by either law alone. Each tells us only what happens in its own domain – that is, how the bodies *would* behave *if* only gravity were present or *if* only

²²¹ Ramsey (2000) suggests this sort of explanation. However, the primary/secondary distinction reflected Locke's attempt to develop an empiricist epistemology that was consistent with the science of his day. We might wonder what might have happened had science then put "shape" in the secondary column. Would Locke have maintained the claim that secondary qualities could not be causal powers?

electromagnetism were present. So neither is strictly speaking true of many actual phenomena, but when amended to be true are less widely explanatory.

Cartwright's point can also be put in terms of properties rather than laws. In the simplest cases, we idealize phenomena as having just one property – e.g. mass but not charge, or charge but not mass – that has its effects independently of other properties. We can then subsume phenomena under laws that deal with only one property. When we explicitly state the *ceteris paribus* clauses, we explicitly acknowledge that the property targeted in the law does not really act independently from other properties (as in the case of mass and charge). However, if the resulting deviation from the behavior that the law predicts isn't significant, the law can be widely explanatory anyway.

My suggestion is that the motivation for seeking intrinsic properties or ignoring extrinsic components of extrinsic properties is the same as the motivation for suppressing *ceteris paribus* clauses. Laws that subsume phenomena under extrinsic properties will have more limited scope, since extrinsic properties are those which can only be instantiated by phenomena in certain types of environments. Acknowledging the extrinsicness of a property is a way of admitting that the nature of the property isn't fully determined by what is internal to a phenomenon. But just as it doesn't follow that the laws (without the clauses) are literally true, it doesn't follow that the properties (without their extrinsic components) are really intrinsic.

Molecular structure may be an example of this. The simplifying idealization here is the rigid-body assumption, in which it is assumed that the distances between the component particles of a complex phenomenon do not vary.²²² When applied to nuclei in

²²² Although literally false, the rigid-body assumption allows us to subsume the motions of composite bodies under our fundamental laws. For example, we can solve the equations of motion for the sun and the

molecules, the rigid-body assumption idealizes away from the internal motions of molecular constituents. Given the state of knowledge when the classical model was developed, there was no reason to think that the model actually simplified nuclear motion. However, the Born-Oppenheimer approximation, proposed in 1927, explicitly acknowledges that nuclei are not literally fixed in position, for when using it we explicitly assume that they are. Its use makes calculating molecular energy much easier and enables us to subsume molecules under chemical laws in virtue of classical structural properties. And, given the classical model, we also had no reason to think that the approximation was not approximately true.

Woolley gives us reason to think not just that the rigid-body assumption is no more true of the nuclei in molecules than it is of macroscopic objects, but that the Born-Oppenheimer approximation is not even approximately true. The results of calculations using the approximation are quite accurate, but that is very different from thinking that the approximation is quite accurate as a description of reality. This is why Woolley attacks the use of the Born-Oppenheimer approximation: its usefulness should not be taken as evidence that it corresponds to an underlying molecular reality.

What the molecular case shows, I think, is how the unexpected, and perhaps unwanted, complexity of a property can remain hidden as an unintended consequence of standard idealizing practices used in the service of explanatory generality. The prejudice against extrinsic causal powers may in turn be an unintended consequence of this standard scientific practice. If Woolley is correct, the particles in a molecule just don't behave in the ways that make classical-type descriptions apply to them unless they are

Earth to an extremely high degree of accuracy without having to solve simultaneous equations for each atom in the two bodies.

also interacting with the environment. If so, then we may just have to accept that molecular structure is a more complex property than we thought. This is like accepting that the forces between two massive bodies just aren't going to be what the law of gravitation says they will be unless we make explicit the assumption (which we know to be literally false) that they are not charged. Similarly, we may have to add *ceteris paribus*-type clauses to our chemical laws to make explicit the assumption (which we would know to be literally false) that the properties cited in the laws are intrinsic.

4.3. WIDE SEMANTIC CAUSAL POWERS

In this section I return to the question of the causal powers of wide content. The section has two parts.

In the first part, I will defend the idea that wide contents may be structural properties of mental representations. First, I use the case of molecular structure to undermine the claim that wide semantic properties cannot play causal roles in a computational model of the mind. We cannot assume *a priori* that all structural properties are intrinsic. So the mere claim that mental representations are manipulated in virtue of their "syntax" cannot bar semantic properties from being robust causal powers. In consequence, the problem of the causal efficacy of wide content turns out to be an artifact of the assumption that a mental representation's structural properties are all intrinsic.

In the second part, I show how semantic structural properties fit into an account of robust and irreducibly mental causal powers proposed by Derk Pereboom. The result is a complete metaphysical account of robust and irreducible semantic causal powers. It remains true that whether contents turn out to be causal powers or not will be a matter of

empirical research. But we will at least know that no general scientific methodological principles or *a priori* arguments bar them from being so.

4.3.1. Semantic Structural Properties

In this section I will argue that the computational model of mind is compatible with wide semantic causal powers. I'll focus on classical computationalism (in its usual "language of thought" form) for two reasons. First, it has the status of being the received view, even though my conclusion is equally compatible with connectionism, the main rival of that view. Second, one of the main issues between classical computationalism and connectionism – the question of whether mental representations have internal structure or not – will itself be recast as a result of my discussion. I will lay out the relevant claims of computationalism before presenting my argument.

The computational model. The central claim of computationalism is that mental processes are mechanical operations defined over the syntax of mental representations. The important components of the model in this context are (1) that syntax, but not semantics, plays a robust causal-explanatory role, and (2) that syntax is intrinsic. Although (1) has been expressed in many ways by many authors, Block gives as clear a statement of it as any:

The problem I have in mind might be put in terms of the 'Paradox of the Causal Efficacy of Content', namely, that the following claims all seem to be true, yet incompatible:

- 1 The intentional content of a thought (or other intentional state) is causally relevant to its behavioral (and other) effects.
- 2 Intentional content reduces to meanings of internal representations.
- 3 Internal processors are sensitive to the 'syntactic forms' of internal representations, not their meanings.

The first claim is meant to be part of the common-sense view of the mind. The third is plausibly taken to be a basic claim of the computer model of the mind, and

the second is a useful and plausible way of thinking how common-sense psychology meshes with the computer model. (1990: 138-139)²²³

Block's 3rd proposition is an expression of (1). The sensitivity of computer – hence brain – mechanisms to the forms of representations but not their representational content is the main claim from which it is inferred that wide contents cannot play causal roles (if the computer model is true). However, the validity of this inference clearly rests on what “form” is thought to be. This leads to (2), the claim that “form” or “syntax” is intrinsic.

A preliminary issue regarding (2) is the ambiguity of “syntax”. This term is used ambiguously to refer both to high-level structural properties of Mentalese sentences, on the assumption that we think in a language of thought, and to low-level but still abstract physical properties (also called “shape,” “form” or “structure”) of mental representations which may or may not be sentence-like. For example, Fodor helped launch the computer model with this influential characterization of “syntax”:

I take it that computational processes are both *symbolic* and *formal*. They are symbolic because they are defined over representations, and they are formal because they apply to representations, in virtue of (roughly) the *syntax* of the representations. ...

What makes syntactic operations a species of formal operations is that being syntactic is a way of *not* being semantic. Formal operations are the ones that are specified without reference to such semantic properties of representations as, for example, truth, reference and meaning. ... ;formal operations apply in terms of the, as it were, “shapes” of the objects in their domains. ...

²²³Fodor appears to express both (1) and (2) in this summary of the classical model:

- i. Thoughts have their causal roles in virtue of, inter alia, their logical form.
- ii. The logical form of a thought supervenes on the syntactic form of the corresponding mental representation.
- iii. Mental processes (including, paradigmatically, thinking) are computations, that is, they are operations defined on the syntax of mental representations, and they are reliably truth preserving in indefinitely many cases. (2000: 18-19)

Here “syntax” seems refer to higher-level properties of sentence-like representations, but it could be read as referring to lower-level physical properties. A connectionist could agree with (i)-(iii) only if “syntax” does not include internal structure; to avoid this possible misunderstanding, she would probably prefer an explicitly lower-level reading of “syntax”.

[S]o long as we are thinking of mental processes as purely computational, the bearing of environmental information upon such processes is exhausted by the formal character of whatever the oracles write on the tape. .. If mental processes are formal, then they have access only to the formal properties of such representations of the environment as the senses provide. Hence they have no access to the *semantic* properties of such representations, including the property of being true, of having referents, or, indeed, the property of being representations *of the environment*. (1980)

We learn from this passage that, at the very least, “formal” or “syntactic” is *not* semantic; it is a case of the intrinsic/extrinsic distinction (see also 1.5.1 for further quotations). This distinction within the computational model has not been challenged.

But “syntax” is often not clearly distinguished from “formal”. At times “syntax” denotes the class of logical-structural properties of sentence-like representations, at other times the class of physical-structural properties of mental representations, whether or not they are sentence-like.²²⁴ These senses are distinguished by Devitt (1986) using the terms “formal” and “syntactic”: formal properties characterize the “shape” of an object, while syntactic properties characterize a structure or system of objects.²²⁵ The formal properties are “fairly brute-physical” or “part brute-physical, part functional” properties, such as being shaped in such-and-such a way or being a certain array of neurons, while syntactic properties are those a linguistic token has in virtue of its role in a language, such as being a noun or being a sentence.

In what follows, I’ll be using the term “syntax” in the lower-level sense (unless explicitly indicated otherwise) to denote physical or physical-functional properties at the

²²⁴ Thus, Fodor (2000) describes computational psychology as involving “an innate Turing architecture of syntactically structured mental representations and syntactically driven computational operations defined on these representations”, this description may be read with either meaning of “syntax” in mind.

²²⁵ Devitt (1986) also describes formal properties (i.e. lower-level syntax) as being intrinsic to particular representations; syntactic properties (i.e. higher-level syntax) are extrinsic to particular representations but not to a system of representations. They are *internal*-relational properties, not *external*-relational. Since I use “extrinsic” in the *external* sense, higher-level syntax would also be intrinsic in the terms I use.

level of implementation – that is, neural properties at some level of abstraction. These are the properties that higher-level syntactic properties supervene on, and that wide semantic properties do not supervene on, if it is true that syntax is intrinsic. These properties are common to all forms of cognitive psychology in which the computational model plays a role. These are also the properties that ultimately matter for robust causal-explanatory purposes. Thus, when Fodor (2000) claims that “cognitive processes are causal only if they are syntactic”, this could mean either that cognitive processes are causal because they are neural, or that they are causal because they supervene on the neural. Wide semantic properties are left out either way, as (1) demands.

Criticism of the model. My criticism of the computational model stems from the analogy that is implicitly drawn in it between syntactic properties and *classical* molecular structure. There are strong parallels between the concept of molecular structure in classical chemistry and the concept of syntax in the classical computational theory of mind. Both properties play essential, robust causal-explanatory roles, and both are (or traditionally have been) assumed to be intrinsic properties. One might say that just as the intrinsicness of molecular structure is part of, or built into, the central dogma of the classical molecular sciences, the intrinsicness of syntax is part of, or built into, the central dogma of the classical computationally-minded cognitive sciences.

Syntax is sometimes modeled explicitly on the classical model of molecular structure. For example, there is a striking similarity between descriptions of syntactic properties and descriptions of molecules in “folk molecular theory”, which is inspired by classical chemistry. In folk chemistry, you take the atoms and glue them together with electron bonds and get something that plays a robust causal role in virtue of those parts

and their arrangement. In cognitive science, you take the representation's parts and put them together (somehow) and get something that plays a robust causal role in virtue of those parts and their arrangement. Thus, Fodor describes syntactic properties as being "constituted entirely by what parts a representation has and how these parts are arranged. You don't, as it were, have to look "outside" a sentence to see what its syntactic structure is, any more than you have to look outside a word to see how it is spelled."²²⁶ Similarly, syntax "reduces to the identity and arrangement of its elementary parts, and, at least in the artificial languages that machines compute in, these elementary parts and arrangements can be exhaustively itemized, and the machine specifically designed to detect them."²²⁷ Here the components of syntax (individual neurons? individual neural impulses?) retain their identities within the larger syntactic unit, just as the atoms retain their identities within the molecule according to the classical molecular model.

A major difference between molecular structure and syntax, however, is that in the former case a whole explanatory framework has grown up around the structure concept, including a causal-explanatory taxonomy based on differences in structure. We have no such broad explanatory basis for syntax – we have incipient neuroscience, and then hand-waving or analogy. This lack of an independent context makes the concept of syntax heavily dependent on the concept of structure elsewhere. Given our interest in the brain – not to mention the numerous thought-experiments in philosophy of mind involving molecular duplicates – the chemical concept is a very likely place to look. This dependence in turn makes the concept of syntactic structure highly vulnerable to the

²²⁶ Fodor (2000: 20-24). If "syntax" in these sentences was understood the higher-level sense, Fodor's claims would be false. You do need to look outside a word to know how it is spelled, because you need to know what it means to know if it is misspelled.

²²⁷ Ibid., 13.

changing fortunes of the chemical structure concept. It is on this basis that we can begin to question the truth of (2): what justifies the claim *now* that all the neural properties lumped under the term “syntax” will be intrinsic? There are three broad reasons why (2) is currently unjustified.

First, to the extent that the assumed intrinsicness of syntax has any *intuitive* justification, it would seem to be derived from the presumed intrinsicness of molecular structure. That is, syntax is intrinsic by analogy to molecular structure. But our best science seems to show that molecular structure isn’t intrinsic. To the contrary, the intrinsicness of structure and the robust causal-explanatory role of structure come apart in a fully quantum mechanical treatment of molecules. That is, the chemical analogues of (1) and (2) come apart; the analogue of (1) can be true even if the analogue of (2) is false. So there is no succor from the physical sciences for computationalists who assume that (1) and (2) must go together in the case of syntax. “Syntax” may still denote that class of neural properties that play causal-explanatory roles in the computational model, but it doesn’t follow that the entire class must consist of intrinsic properties.

Second, there is no empirical justification for the assumption that syntax is intrinsic. Our current state of knowledge of the cognitive architecture of the brain can’t possibly justify it. At this time, we can make only fairly gross distinctions between cognitive-functional kinds: visual processing (for action or perception), phonetic processing, face recognition, short-term memory storage and so forth. We know to a greater or lesser extent which neural networks (and in some cases, single cells) are most active for broadly distinct kinds of stimuli. But we have little to no empirical detail that would enable us to distinguish particular types of cognitive processes or states, let alone

their tokens. We don't know what makes some particular type of firing pattern a face-representation or a memory-trace or a water-belief, let alone whether some token firing pattern is a Mom-face or a Dad-memory or a token belief that water is wet. So it is only a leap of faith to declare that the syntactic properties of mental representations are intrinsic. For if we don't know anything about the details of mental representations, we can't know what their structural properties are, and we certainly can't know that they are intrinsic.

This lack of empirical detail becomes painfully obvious whenever syntax is described. The descriptions are invariably couched in terms appropriate to the higher-level sense of syntax, and it is never made clear how the analogy to neural or physical properties is supposed to go. Apparently, whatever properties there are that are consistent with materialism and mechanism will do. But this of course would not rule out extrinsic properties. It is consistent with the computational story that environmental differences necessitate local differences, and in particular that environmental differences determine causally relevant differences among structural properties of mental representations. If they make a difference to molecules, there is no *a priori* reason why they cannot also make a difference to mental representations.

In short, the taxonomy of structural properties (syntax) in cognition is completely wide open. Since we simply do not know which local microstructural properties play important roles in cognition, we don't know what their nature is either. Declaring now that syntax is intrinsic is armchair speculation of the boldest and baldest sort. Moreover, if current taxonomic practice in cognitive neuroscience is any indication, it is extremely likely that at least some syntactic properties will be individuated partly by their cognitive-functional role. So we have every reason to think that a cognitive taxonomy will be

extrinsic at least in part, and no reason to think that any hypothesized reduction of syntactic properties to even lower-level properties will somehow usher in intrinsicness.

Third, there is also a *prima facie* difficulty for any account of the physical nature of mental representations in which the syntax of the representations is considered entirely intrinsic. This is the fact, already noted in Ch.3, that the brain exhibits great plasticity during an individual's development and, to an extent not yet known, over the individual's lifetime. I argued that plasticity motivates extrinsic individuation of local microstructure because it forces us to identify significant differences in microstructure partly on the basis of the cognitive and behavioral differences that result. The latter, of course, will often be individuated partly in terms of what is the case in the agent's environment.

The applicability of this argument to the case of syntax (rather than local microstructure) requires identifying syntax with that subset of local microstructure that is involved in cognitive activity. However, this identification seems justified. There seems to be no reason to think that philosophers of mind intend to refer to different things when they use the terms "syntax" and "local microstructure". (Higher-level syntax is a different story, as are neural properties that are not involved in cognition.) Both terms are used to refer to the set of largely unknown physical-neural properties that play implementing roles in cognition. And since we are never told in appropriate terms what syntax is, the claim that it is *not* identical to local microstructure would be a bare assertion. We do have a gross taxonomy of local microstructure, though, and it seems odd to deny that at least some syntactic properties are local microstructural properties, at this or a more fine-grained level. But if syntax is, or is a subtype of, local microstructure, then the *prima*

facie difficulties raised in Ch. 3 for the intrinsic individuation of local microstructure will apply equally to the case of syntax.

It should be noted that giving up (2) – the claim that syntax is intrinsic – has no negative repercussions for the computational model as such. The intrinsicness of syntax is a customary assumption, but it is not an essential element of the model. The claim that mental representations are manipulated mechanically in virtue of their syntax is a form of hand-waving in the direction of causal interactions between mental particulars in virtue of physical properties that are analogous to (or maybe are) shapes, forms or structures. Higher-level syntactic properties are implemented in the lower-level syntactic properties. Whatever the relation might be between the two levels of syntax, the typical broad claims about the causal role of syntax at either level do not illuminate or require us to assume anything about the nature of lower-level syntax other than that it must be accessible to mechanistic manipulation by a machine. These claims are invariably packaged with the assumption that syntax is intrinsic. But this unjustified claim about the nature of syntax can be abandoned without harming the computational model at all.

In sum, the analogy to chemistry fails to provide support for the claim that syntax is intrinsic, and our empirical evidence is far too sketchy to enable us to determine the nature of any syntactic property, let alone all of them. The remaining question is why we might think that some of the structural properties now called “syntax” might be classified as semantic. Ultimately whether there are semantic structural properties will be an empirical matter. But in theory the essential dependence of wide semantic properties on the way things are in the speaker’s environment is fully compatible with the claim that these properties are encoded or implemented directly in structural properties of mental

representations, just as we presume the higher-level syntactic properties are. There is no reason to think that semantic properties will be left out of a fully computational story. *Structural* properties of some sort – the shapes, forms or structures we now denote using the term “syntax” – are essential to the model, but the assumption that these properties are all *intrinsic* is not. The model demands only that mental representations be manipulated in virtue of their “syntax”, but it does not demand that all the properties now lumped under the term “syntax” be intrinsic. It allows that at least some of these structural properties can essentially depend for their individuation in part on environmental influences on the individual, both in the course of development and over the individual’s lifetime.

Implications. In this section I’ll draw out some implications of the above discussion regarding the problem of externalism, the nature of narrow content and the debate between classical computationalism and connectionism.

1. Regarding the problem of externalism: If I’m right, then since “syntax” is already considered robustly causal, I’ve also shown how it is possible that wide contents are robust causal powers. On this account, we have simply used the label “syntax” to label a set of structural properties that we’ve assumed are all intrinsic, but which actually includes wide semantic properties (or their subvenience bases). The problem of externalism thus turns out to be an artifact of the assumption that the structural properties of mental representations are intrinsic. This assumption creates a theoretical situation whereby we need to explain the *prima facie* causal role of wide content in Rube-Goldberg fashion, or explain it away entirely. We must either get semantics and syntax to covary somehow, or we must find or invent a ‘narrow’ version of content, or we must

give up on semantics. The problem dissolves if the assumption of the intrinsicness of structure is abandoned.

We saw in the case of molecular structure that the possibility of extrinsic structural properties does not impugn the causal-explanatory value of these properties or the scientific status of the theory in which they play a causal role. By the same token, the causal properties of mental representations need not differ if the structural properties we use to explain their behavior turn out to be determined in part by their relations to the environment. Nothing essential to the computational story needs to change if some structural properties of mental representations are not intrinsic. That assumption is entirely gratuitous to cognitive psychology, and may be no more fruitful for future research in cognition than the assumptions of classical chemistry.

2. Regarding narrow content: If the extrinsic hypothesis is correct, the nature of narrow content would be slightly altered. Narrow content would be a determinable state or process – a state or process that could be a “cat” or “gato” or “chat” or (etc.) ... in the right environment. This is similar enough to Fodor’s conception of narrow content as a function from context to wide content, one in which narrow content is not expressible. But narrow content would not be the causal core of wide content. A given narrow content would be like an isolated isomer – a state without a determinate identity or causal powers until its nature is determined by environmental interaction. In other words, narrow contents would be states of an isolated brain, not states of an embedded brain.

3. Regarding the debate between connectionists and classical computationalists: One of the main issues that divide the parties is whether mental representations have internal structure or not. This debate and its resolution would also be slightly recast.

Connectionists, who largely deny that mental representations have sentence-like structure, would be correct in the case of isolated or deprived brains – neural states in these brains would not play determinate causal roles in cognition. One would study these non-structured isolated states (in theory if not practice) in order to understand the role that the internal components play in cognition. However, classical computationalism would be correct for the neural states in normal brains, which would have the necessary environmental interactions to determine the identity and causal roles of the states.

This suggests a different way of reconciling the difference between the structured mental representations of classical computationalism and the unstructured mental representations of connectionism. Fodor and Pylyshyn, for example, have suggested that connectionism may be correct at the level of implementation; the neural networks of connectionism would implement classical computational states. It is not said how this resolution is supposed to go, especially since connectionism is supposed to be a *cognitive* theory. On the view suggested, the reconciliation is much simpler. It would be a matter of whether the mental states are being studied in isolation or not. Thus, classical and the connectionist stories can be on a par in terms of cognitive levels of explanation and their relation to neurological or neurophysiological levels of explanation; they differ, albeit compatibly, in that they explain cognition in different environmental contexts.

4.3.2. Robust Mental Causal Powers

In this section I'll show how the above account of semantic structural properties coheres with Pereboom's (2002) defense of robust nonreductive physicalism, which includes a defense of irreducibly mental causal powers as its key feature. What his

account adds is the idea that some mental causal powers are *irreducibly* mental. This is a good further feature to have, since it would show that the structural properties that might be classified as wide semantic are not necessarily reducible to something else. Since Pereboom (2002) builds on ideas expressed in Pereboom (1991) and Pereboom and Kornblith (1991), the following exposition draws on all three papers. I will save my comments until after it has been presented.

The Token-Level Account: Constituted Tokens. At the token level, Pereboom denies the identity of mental tokens with physical tokens and token-level mental causal powers with the causal powers of physical tokens. Token causal powers are the powers that a token may exercise in a singular causal interaction in virtue of its instantiating a given property at that time. For example, the chemical state of a sleeping pill at t causes John's drowsiness at $t+1$. The claim is that the mental causal powers of a token are constituted by, not identical to, those of its microphysical realizer on a given occasion of instantiation:

The causal powers of a token of kind F are constituted of the causal powers of a token of kind G just in case the token of kind F has the causal powers it does in virtue of its being constituted of a token of kind G . (1991: 131), (2002: 504).

Pereboom uses the ship of Theseus argument – familiar from debates about material constitution – to argue for the irreducibility (non-identity) of both token mental causal powers and the mental tokens that instantiate these causal powers.²²⁸ The original argument holds that the ship of Theseus is not identical to its current token microphysical realization because it could have had, or could have, a few different planks and it would have remained, or would remain, the same ship. There can be two distinct objects – the

²²⁸ See Rea (1995) for an overview, as well as Baker (1997) and Johnston (1992). Hornsby (1980) and Horgan and Tye (1985) provide other arguments against token-token identity.

ship and its physical “constitutor” – that coincide in their current material constitution but differ in their temporally extrinsic or modal properties.

Similarly, my token desire for ice cream, M1, is not token-identical to its actual microphysical realizer P1. For if the event M1 had had the slightly different event P2 as its microphysical realizer rather than P1, it still would have been M1. Its identity would not have changed although its constitution would have. The argument is then extended to the case of token causal powers:

Is the token causal power of your present desire for ice-cream, D, identical to the token physical causal power, P, which constitutes it? No. Suppose that P is a token causal power of the molecules, M, that constitute a token brain cell, and that you ingested these molecules while eating your favorite baby cereal on the morning of your first birthday. Suppose also that your mother had the choice of two type-identical boxes of this baby cereal on the previous day, and that she chose the one on the left. If she had chosen the box on the right, you would not have had token molecules M (although you would have had molecules M' of the same type), and because you would lack these token molecules, you would also lack their attendant token molecular causal power P (although M' would have had a molecular causal power of the same type). Nevertheless, you clearly would have had the very same token desire for ice-cream with its token psychological causal power D. Hence, a token psychological state and its token psychological causal power can remain the same even when its token molecular constitution, and thus its token molecular causal powers, are altered. (1991: 132)²²⁹

Thus, my current desire for ice cream causes my going to the corner ice-cream store. The token desire is constituted by but not identical to a microphysical state, and its causal powers are constituted by but not identical to the causal powers of that microphysical state. By analogy, the causal powers of Theseus's ship would not be identical to the causal powers of the composite that realizes it at any given time, for the ship's causal powers would stay the same if we replaced a few planks. In general, multiple realizability

²²⁹ At (2002: 500) he summarizes: “The token mental causal powers of M1 and M2 will not be identical with the token microphysical causal powers of P1 and P2, nor with the token neural causal powers of the neural states N1 and N2 that constitute M1 and M2, nor with token causal powers at any other level of description more basic than the neural.”

at the type level is evidence that the token causal powers invoked in a higher-level explanation are not identical to those in an explanation given by a lower-level science.

The irreducibility of token mental causal powers is what makes them robust. This claim of irreducibility is quite different from the claim that some of the token M's causal powers are classified under a mental type that is not reducible or identical to a physical type. For it is easy to classify the *same* causal powers in different ways under non-identical descriptions. This possibility drives the reductionist (and "proleptic") claim that lower- and higher-level explanations use descriptions that pick out the same causal powers – namely, lower-level properties. As Pereboom notes, however, if type-type identity is false but a token-identity thesis is true for M1's causal powers,

then the causal powers to which the psychological explanation refers would in the last analysis, in fact, be microphysical. Psychological explanations might then presume a classification that clusters microphysical causal powers in a way distinct from how microphysics sorts them, but this would not compromise the microphysical status of those causal powers. Hence, robust nonreductive materialism affirms various token-diversity claims for mental causal powers. (2002: 500)

If psychological tokens and their causal powers were identical to some composite or "cluster" of microphysical tokens and their causal powers, then the clusters would have no ontological or robust causal significance. They would merely cross-classify the basic tokens and their causal powers for higher-level explanatory reasons.

The Type-Level Account: Intrinsic Mental Types. At the type-level, Pereboom argues that there can be intrinsic yet irreducibly mental properties that are the causally-efficacious realizers of functionally defined mental types.²³⁰ The motivation for this view

²³⁰ Pereboom follows the traditional metaphysical distinction between universals that are types or kinds from universals that are properties; kinds have (or are constituted by) properties. So, for example, the kind human being has (or is partly constituted by) the property of having a particular genetic makeup; the kind pain has the (dispositional) property of causing pain-avoidance behavior; the kind diamond has the property

is Kim's argument that the causal powers of a functional type just are the causal powers of its lower-level realizer on a given occasion of realization.²³¹ According to Kim, just as there may be species or structure-specific reductions for multiply realizable mental types, the causal powers of the higher-level type are similarly reducible to (can be identified with) those of its token realizers. For, Kim claims, if a system has a functional property F on a given occasion in virtue of its having realizer Q, then "it is difficult to see how we could avoid saying" that the causal powers of this instance of F are identical with the causal powers of this instance of Q. So to the extent that the functional property has causal powers, they are nothing more than the causal powers of its realizer on a given occasion of instantiation. Pereboom calls this claim Kim's "causal inheritance principle". If Kim is right, then while functionalism preserves the autonomy and irreducibility of mental types, it does so at the cost of giving up their claim to robust causal power status.

Pereboom agrees that the realizers of functional types are the genuine causal powers, so he agrees with Kim's conclusion about mental types. He also goes along with Kim's assumption that these realizers are *intrinsic*, even though Kim's "causal inheritance principle" is in principle compatible with the claim that the realizers of functional types are extrinsic properties. But *pace* Kim's implicit assumption about the nature of these realizers, Pereboom claims that the causally-efficacious realizers of mental types can be *mental* properties. He neither addresses nor rules out the possibility

of having a certain chemical structure. I'll remain neutral on this metaphysical issue and just use the term "property" indifferently when stating his position. Where needed, the "first-order" vs. "second-order" distinction will do.

²³¹ See Kim (1992/1993). In Pereboom's words, functionalism "identifies mental states with type-level dispositions to cause mental state and behavioral outputs given perceptual inputs and mental states – with the understanding that these dispositions are purely relational, that they be analyzed in terms of causal relations to perceptual inputs, behavioral outputs, and other mental states, and no intrinsic components." (2002: 511)

of defending robust yet extrinsic mental causal powers in the form of extrinsic realizers.²³²

Pereboom uses two analogies – one explanatory, one metaphysical – to argue for the claim that it is plausible to posit intrinsic, irreducibly mental realizers. The first analogy extends to psychology a general scientific explanatory strategy by which dispositional features are explained by non-dispositional properties within the same science.²³³ For example, the dispositional features of chemical compounds – such as their combinatorial characteristics – are explained by their chemical structure²³⁴; and certain features of diseases – such as their contagiousness – are explained by intrinsic biological properties. Functionalism defines mental types in terms of their dispositions to cause other mental states and behavioral outputs and to be caused by environmental inputs and other mental states. So, by the explanatory analogy, it is plausible to suppose that there may be intrinsic mental properties that explain the dispositional features characteristic of mental types.

What these intrinsic properties might be is largely left open. He suggests that they might include phenomenal content, and also intentional properties: even if externalism is true, “it remains natural to suppose that the content of a belief type is at least partially

²³² Nevertheless, I doubt Pereboom would endorse extrinsic realizers. He considers (p. 351-2) the functionalist who suggests that the underlying genetic or molecular properties that explain functionally-defined biological kinds are themselves functional or non-dispositional, and thus only apparently intrinsic. He responds that this possibility doesn't rule out that they have intrinsic properties too. True enough; but the point remains that even if there are intrinsic realizers, nothing rules out the causal efficacy of extrinsic realizers. (Of course, the molecular properties themselves might be extrinsic.)

²³³ Pereboom ascribes this general strategy to Putnam (1975), Boyd (1999) and Van Gulick (1995).

²³⁴ Pereboom seems to assume intrinsic molecular structure. I am simply presenting his view at this point.

intrinsic to it.”²³⁵ But the second, metaphysical, analogy he uses to argue for irreducibly mental intrinsic properties provides a clue as to what kind of property they might be:

Consider the analogy of a ball-piston engine, ... which has a specific internal structural configuration. Characteristic of this engine is its having parts with particular shapes and rigidities, and these parts must be arranged in a particular way. These features are manifestly not functional relations that such an engine stands in; rather, they constitute intrinsic characteristics of this type of engine. At the same time, these characteristics are multiply realizable. The parts of the engine can be made of material of different sorts – as long as the material can yield, for example, the required shapes and rigidities. The ball-piston engine, then, has a nonfunctionalist intrinsic structure that instantiates its causal powers, but it nevertheless admits of distinct realizations. By analogy, it might be that the heterogeneous physical realizations of the dog’s and the human’s belief that there is danger nearby exhibit a structure of a single type that is intrinsic to this kind of mental state, a structure that instantiates the causal powers of this belief. This structure may be more abstract than any specific sort of neural structure, given that it can be realized in distinct sorts of neural systems. (2002: 517-8)

So just as the ball-piston engine’s structure is intrinsic and multiply realizable, there can be intrinsic and irreducible mental properties. “Being an engine” is a functional type multiply realizable in different intrinsic structural properties – ball-piston, rotary, hydrogen-burning, hybrid; in turn, each of these structural properties is multiply realizable in different materials – steel, aluminum, titanium, plastic. Similarly, a given functional mental type may be multiply realized by intrinsic structural mental properties. These properties would be intrinsic because they are structural, and they would be irreducibly mental because structural properties can be realized in different types of physical matter.²³⁶

²³⁵ Pereboom holds that the standard thought-experiments for externalism “fail to show that *nothing* in [a subject’s] belief contents remains the same, so it does not show that every aspect of having the belief with that particular content is an extrinsic property of her.” (1991: 351) This is of course compatible with my argument against Fodor that the standard-thought experiments do not show us anything about the relation between inner states and the environment. My position entails his.

²³⁶ See also Shapiro (2000) on multiple realizability.

Assessment. Clearly, if structural properties are analogous to the irreducibly mental, robust causal powers that Pereboom has in mind, then extrinsic, semantic structural properties would be irreducibly mental, robust causal powers as well. There is nothing in this account that precludes this possibility. Moreover, as we saw above with the computational model, the assumption of the intrinsicness of structure appears as a gratuitous and inessential element of the model. There is simply no need to make that assumption. So I will not repeat the arguments I gave above for that claim, and instead will close this section with two brief comments to defend Pereboom's account against two possible objections.

(1) The type- and token-level accounts work together to support robust mental causal powers, even though they are logically independent of each other. (One could affirm both token-token identity and intrinsic mental realizers, or deny both token-token identity and intrinsic mental realizers. Mainstream nonreductivists affirm token-token identity but deny intrinsic mental realizers.) The way they work together can be seen by imagining a certain line of counterargument. According to Pereboom, the explanatory-strategy analogy supports the positing of intrinsic properties to explain the functional ones, and the ball-piston-engine analogy supports the positing of structural properties, which are multiply realizable, as the sort of intrinsic properties that might do the job. I would modify this only to say that the explanatory strategy analogy supports the positing of "local" properties and the engine analogy supports the positing of structural properties, thus leaving out the gratuitous claims about intrinsicness.

Both analogies are motivated by nonreductivism. But the explanatory-strategy analogy is tailor-made for reductivism, as Kim's longstanding support for structure-

specific reductions should make clear. Thus, even if Pereboom is correct that these structural properties are multiply realizable, the counterargument is that we can identify each instance of a structural property with its material realization in silicon, carbon, steel, and so on. Where Pereboom posits an irreducible intrinsic mental property to realize functional mental types, a reductivist might argue that the structural properties can be identified with their microphysical bases on a given occasion of instantiation.

This is where the token-level constitution account plays an essential role. For Pereboom can reply that each structural property instance cannot be identified with its constitution, since he has denied that constitution is identity at the token level.

The same dialectic goes for the defense of robust mental causal powers as well. If this objection were right, Pereboom's defense of robust mental causal powers would be to no avail. But again, the token-level account comes to the rescue, for it enables Pereboom to propose a "causal inheritance principle" that replaces Kim's reductive version:

If mental property M is realized in a system at t in virtue of physical realization base P, the causal powers of this instance of M are wholly constituted by the causal powers of P.

So we can summarize the two accounts as follows. If F is a functional mental type, it can be realized by an intrinsic mental realizer M, and the causal powers of this instance of F will be those of M. But M and its causal powers are constituted in, but not identical to, its physical base P and P's causal powers. So while the functional mental type is not causally efficacious, the causal powers of M remain mental. They reside in the realizer of F, as Kim would agree, but *pace* Kim the realizer is itself a mental property. What I would modify in this account is just the assumption that the mental realizer is intrinsic. The fact that it is structural does not entail this feature.

(2) It might be thought that an argument against the identity of mental and microphysical *tokens* – whether the ship of Theseus or some other argument – is not sufficient to show the failure of identity between *token causal powers*. Suppose it is true that my token desire for ice cream could have had a slightly different microphysical makeup without this difference making any difference in the identity of my desire. It would still be a token of that type. But does it also make sense to say that a token's causal powers would not differ if its molecular constitution differed? In the case of causal powers, “would *be* the same as” is simply “would *do* the same as”. It may be questioned whether a token desire D actually realized by P would *do* the same as D would have had it been realized by the slightly different P'. Isn't it plausible that D would *do* (or be capable of doing) something slightly different?

The question may be posed by comparing the ship of Theseus argument and Fodor's Non-Locality argument. Fodor claimed that a difference in content can make a causal difference only if it makes a local difference; causal power differences are individuated (or identified) with local differences. Pereboom claims that local differences do not entail causal differences; causal-power differences are individuated (or identified) with content differences. Each side musters persuasive intuitions in its favor. Surely, says Fodor, if *atoms* don't differ, causal powers don't differ; surely, says Pereboom, if *mere* atoms differ, causal powers don't differ.

I think Pereboom wins, but not because the ship of Theseus is such a good argument. I argued in Ch. 3 that causal power differences can be individuated with local differences, but local differences – in particular, local microstructural differences – aren't the same as *intrinsic* differences. For if local microstructure is individuated extrinsically,

then causal powers will be individuated extrinsically. Pereboom rejects classification by “local” differences on intuitive ship-of-Theseus grounds, with his desire-for-ice-cream example. What possible difference could a teeny weeny atom make? Well, it could make a lot of difference. But that’s not relevant. For science shows that causal powers can be individuated when there is no “local” difference (as with chiral molecules) and they can be individuated with atomically “local” yet extrinsic microstructural differences (as with structured molecules in a quantum mechanical treatment). In both cases, causal powers are individuated with extrinsic differences, whatever the case may be with the atoms.

In other words, where there is extrinsic individuation, the idea that causal powers needn’t differ when atoms do will be correct. But our intuitions about mere atomic differences are symptoms of this fact. That is, it is because the ship of Theseus is individuated extrinsically that the “mere atoms” don’t matter; it is not because “mere atoms” don’t matter that the ship is individuated extrinsically. Just so, cognitive neuroscience seeks taxonomic categories that will cover all normal human brains as well as the brains of at least some other species that brain areas. This motivates individuating brain areas in part by reference to cognitive function. It is a consequence that “mere atoms” won’t make a taxonomic difference – not because they’re mere atoms, but because we don’t expect a difference in atoms to make a difference in cognitive function. It will *remain true* even on this story that causal powers are entirely *constituted of* microphysical states – but that’s exactly what Pereboom’s constitutional account buys.

4.3. CHAPTER AND DISSERTATION SUMMARY

In this chapter I presented an account of how extrinsic properties might be robust causal powers.

First, I used the case of molecular structure to show that at least some properties that are considered robust causal powers can be extrinsic. I first introduced the debate in theoretical chemistry about molecular structure, in particular the claim that the classical model that is deeply ingrained in popular intuition is false. Although the molecular structure debate is not over, at the very least the intuitive molecular model can no longer be taken for granted. I also suggested that the resistance to extrinsic causal powers may be explained as an unintended consequence of the standard scientific practice of idealizing away from the complexities of actual phenomena.

Consequently, I explored the implications of an extrinsic view of molecular structure for the debate about wide content. I argued that the claim that semantic properties can't be causal powers in the computational model rests on analogies to the classical view of molecular structure. If this is right, then the claim that semantic properties can't be causal powers rests on a mistake. The mistake is to think that the class of properties now labeled "syntax", which is granted to be causal-explanatory, is entirely intrinsic. This assumption about the nature of syntax is unjustified.

I then showed how Pereboom's defense of irreducibly mental robust causal powers, which denies that constitution is identity for mental tokens and their token causal powers, can be extended to include wide semantic properties if these can be structural properties. If so, then we have a complete metaphysical account of how wide semantic properties can be robust causal powers. This positive view would complement the

responses to the arguments presented in Chapters 2 and 3 that aimed to show why they cannot be robust causal powers.

At this point the jury is out; we will have to see what empirical research reveals. But there is no reason in principle why the causal laws in a scientific psychology cannot subsume phenomena in virtue of their wide semantic properties. To the contrary, the door is wide open to a scientifically respectable psychology in which wide semantic properties play a robust causal role.

APPENDIX I: WOOLLEY'S ARGUMENT IN MORE TECHNICAL TERMS

A somewhat more technical version of Woolley's (1978) argument is as follows:

- 1) Calculations of molecular energy from basic quantum principles do not yield distinctions in the energy of identical nuclei.
- 2) Chemical experiments with molecules reveal distinctions in the energy of identical nuclei that are reliably interpreted in terms of classical structures.
- 3) These interpretations of our experimental results in terms of classical structures are not consistent with quantum mechanics.
- 4) Experimental results in chemistry should be interpreted consistently with quantum mechanics.
- 5) So we should abandon the classical model as false.

I'll explain each of these premises in detail.

Premise 1. *Calculations of molecular energy from basic quantum principles do not yield distinctions in the energy of identical nuclei.*

Woolley takes as his starting point the calculation of the *complete* and *exact* energy of an *isolated* molecule from basic quantum principles, without making any of the usual simplifying assumptions used by chemists (see below).²³⁷ An *isolated* system is one that is not interacting – not exchanging energy or mass – with its environment.²³⁸ The

²³⁷ In quantum mechanics, each measured value of an observable (energy, linear momentum, etc.) has associated with it a function (an eigenfunction) which expresses the probability of getting that observed value (an eigenvalue) at a given instant. Each type of observable is represented by an operator such that, when a given measurement is made, the obtained value is an eigenvalue of a particular eigenfunction (or eigenstate) operated on by that operator. In philosophical terms, an eigenfunction (eigenstate) is a dispositional property of a system: System M is in eigenstate S at t = if M were measured at t, M would (with probability P) be observed at t to have an instance of the observable-type S with value v. The general form of an eigenvalue equation is $\hat{A}\varphi = a\varphi$, where $\hat{A}$ is the operator, φ the eigenfunction and a the observed eigenvalue. Information about a particular observable – e.g. the probability of observing a particular linear momentum when we next measure it – is obtained by solving an eigenvalue equation for that observable. When we want to calculate an isolated system's energy at an instant, we use the Hamiltonian operator $\hat{H}$, which corresponds to the total (kinetic and potential) energy of the system in terms of coordinates and momenta, and solve the time-independent Schrödinger equation, $\hat{H}\psi = E\psi$, for that system. If at any time a system is in a state for which this equation has a solution, the probabilities of observing particular values of the system's observables are constant or non-variable over time. For this reason the eigenfunctions (or eigenstates) of the Hamiltonian are also called stationary or time-independent states.

²³⁸ A closed system is one exchanges energy but not mass, and an open system is one that exchanges both. Molecules in natural conditions are closed systems; this is why molecular weight can be intrinsic even if molecular structure is not. Woolley (1985: 1083) defends the isolation assumption as an idealization on a par with the role of the principle of inertia in mechanics, as when Galileo assumed a body which moves without the influence of external forces in order to develop his new science of dynamics. The isolation

energy of such a system is constant over time due to the principle of conservation of energy. For this reason an isolated molecule is also said to be in a *stationary* or *time-independent* state, in that as the system evolves in time the expected values for its energy are a constant (in time) function of its initial energy state. Isolation also allows us to calculate the molecule's *exact* energy. In contrast, a system that is *not* isolated is exchanging energy with its environment. Energy exchanges take time. So due to the complementarity of time and energy, it is impossible in times less than infinity to specify exactly the expected values for the energy of a non-isolated system; the best we can do is specify an expected range of values, a range that will differ in time. For this reason a non-isolated system is said to be in a *transitional* or *time-dependent* state. Lastly, a *complete* calculation involves using separate terms for each of the types of energy of each of the constituent particles, nuclei and electrons alike. This means each particle contributes to the total energy of the system on equal footing in the sense that we do not ignore any interactions among the particles which can affect their motions and hence their energy.

In these calculations of molecular energy, the distinctions of energy between identical nuclei that they would have if the classical model were correct do not appear. The point requires a bit of explanation. In a classical model, nuclei that are fixed in different positions in the molecule will exhibit differences in their energy. For example, a centrally located carbon nucleus will have smaller angular momentum than a peripheral carbon nucleus. As Woolley (1978: 1077) notes, "[i]t is characteristic of the [classical] molecular structure description that identical nuclei can be dynamically inequivalent."

assumption is also widely used in physics and chemistry textbooks, and conditions of virtual isolation are nearly achievable experimentally.

But in quantum mechanics (unlike classical mechanics) identical particles – two carbon nuclei, for example, or two electrons – are *indistinguishable*: there can be no observable difference in the physical properties of a system if identical particles are exchanged.²³⁹ They are fully *symmetric* – completely interchangeable – under various operations, including rotation, reflection and permutation (exchange of position). This means that all identical nuclei in an isolated molecule are *precluded* from contributing different amounts of energy to the total energy of the system. In effect, identical nuclei in an isolated molecule are just as “delocalized” as electrons – they have no fixed position in the molecule. It is impossible to measure the position of *particular* identical nuclei.²⁴⁰

In terms of presenting Woolley’s argument, this is sufficient for understanding the first premise. For the record, however, the simplifying assumptions that Woolley omits are those chemists nearly always employ in order to make their calculations more tractable, given that molecules are many-body systems (the smallest, H₂, has two nuclei and one electron). These assumptions simplify the interactions among the constituent particles by ignoring some and calculating only the remainder.²⁴¹ For example, we may ignore electronic and nuclear interactions by assuming an adiabatic separation of their motions; we assume their motions do not involve any exchange of energy. (Woolley’s calculation procedure is non-adiabatic.) Foremost among these assumptions is the Born-

²³⁹ Liboff (2003: 597) remarks: “At the quantum mechanical level of description, identical particles are also *indistinguishable*. In the classical description of a system of identical particles, one may conceptually label such particles and follow their respective motion. This is impossible at a quantum level. There is no experimental result that distinguishes between two states [wavefunctions] obtained by exchange (interchange) of identical particles.”

²⁴⁰ Claverie and Diner (1980: 58) remark: “For example, rotational symmetry implies that the system of nuclei should be described as a “nuclear cloud” (exhibiting spherical symmetry), as well as an electronic cloud, for those molecular eigenstates – the ground state among them – which belong to the identity representation of the rotation group.”

²⁴¹ The following non-technical explanation of the many-body problem is instructive: “Many-body problem may be defined as the study of the effects of interaction between bodies on the behaviour of a many-body system. One of the preferred methods for solving the problem is simply to ignore it, i.e., pretend there is [sic] no interactions present.” See <http://explanation-guide.info/meaning/Many-body-problem.html>.

Oppenheimer approximation, in which the nuclei are assumed fixed in position – that is, their kinetic energy is set to zero, so their energy is all potential – and we solve for the electronic motions as if they were not affected by nuclear motion.²⁴² The resulting calculations are quite accurate. This is because the nuclei are so much more massive than the electrons and so move much more slowly relative to them; compared to the electrons, they *are* fixed in position. The approximation also reinforces the classical model since by using it we treat nuclei as quasi-classical particles.²⁴³ By *not* making this or related assumptions, Woolley treats the nuclei as fully quantum mechanical particles.

Premise 2. Chemical experiments with molecules reveal distinctions in the energy of identical nuclei that are reliably interpreted in terms of classical structures.

Observations of molecules and molecular energy levels are not direct.²⁴⁴ We infer to classical molecular structures from spectroscopic and other indirect measurements of molecular energy. In spectroscopy, prepared atoms or molecules (e.g. in a low-pressure, low-temperature gas) are irradiated with electromagnetic radiation in a known range of frequencies. In this interaction, targeted molecules undergo transitions to new energy levels, absorbing and emitting photons (quanta or particle-like packets of electromagnetic energy) of different frequencies or energy levels. For example, a nucleus will vibrate

²⁴² More precisely, we make a series of approximations of the internuclear distances and solve for the electronic motions for each; this gives us a nuclear potential energy curve (since the nuclear energy is all potential); a minimum on the curve corresponds to an equilibrium bond length. A more complex procedure yields a potential energy surface, which also will yield equilibrium bond angles. See Atkins.

²⁴³ Weininger (1984: 940) observes that “by adopting the fixed nucleus approximation quantum chemists guarantee that they will achieve results that are interpretable in terms of one or more molecular structures.” He also concludes: “In order to derive classical molecular properties theoreticians have always found it necessary to introduce classical molecular constraints into their calculations.” Regarding the justification of the assumption to chemistry students, he remarks: “The characteristic reaction of students to the introduction of these assumptions is twofold – they are impressed by the resulting simplification and puzzled as to the origin of the assumption. “Why did you assume that?” is a typical question. The answer is of course that we have a strong prior expectation of what nature *should* look like, which strongly conditions our notion of what is “physically reasonable”.”

²⁴⁴ Claverie and Diner (op. cit., 65) note that no one knows how to perform a *direct* measurement of molecular energy, let alone observe molecular structure.

more vigorously or an electron will “occupy” a higher-level orbital. A particle that moves to a higher energy state will absorb energy at a certain frequency; one that moves to a lower energy state will emit a photon of a certain frequency.²⁴⁵

The frequencies of the emitted or absorbed photons are measured and recorded as peaks or lines on the spectrum. (In absorption spectroscopy, beams are passed through sample and reference cells and the relative absorption is measured.) Different locations of lines on the spectrum can be associated with different levels of kinetic energy or motion; for example, vibrational transitions are associated with photons with wavelengths in the far infrared range (0.03 mm), while rotational transitions are usually associated with microwaves (3 mm). The spectral lines or patterns are then correlated with differences in bond strengths, lengths and angles, which allows us to infer the type of classical structure

²⁴⁵In spectroscopy we analyze the emission or absorption spectrum to determine the energy levels in the molecule and from there infer to classical structures. The classical idea behind the procedure is that for the molecule to interact with an electromagnetic field and absorb or create a photon of a given frequency, it must possess, at least momentarily, a dipole (a two-charge array of point charges with no net charge, which corresponds to a classical bond) oscillating at that frequency. (Atkins (1998: 461). Intuitively, when an object – a guitar string, a dipole – vibrates at a particular frequency, it will absorb energy from a vibration of exactly the same frequency. In IR (infrared) spectroscopy, for example, we take advantage of the fact that infrared light happens to be in the same frequency range as the vibrations in molecules. See www.psrc.usm.edu/macrog/ for a non-technical explanation.

The frequency of a wave is how many times per second its displacement at a fixed point returns to its original value. The wavelength is the distance between neighboring peaks in a wave; shorter wavelengths have higher frequency, hence higher energy (in particular, higher momentum, according to the de Broglie equation relating wavelength and momentum). For example, visible red has a longer wavelength and lower frequency than visible violet. A wave's wavenumber is the number of complete wavelengths in a given length (e.g. a wavenumber of 5 cm^{-1} means that there are 5 complete wavelengths in 1 cm.). The wavelength is the wavenumber $\times$ the speed of light. Technically, in spectroscopy we detect wavenumbers, convert them into wavelengths and plot the wavelengths on a spectrum.

Spectral lines vary in width. The less precise the lines, the less precise the energy levels in the molecule can be specified; equivalently, the lower the resolution of our measuring device. This is where time-energy complementarity enters the picture (see also fn. 24 below). The shorter the time that the system is interacting with the measuring instrument – that is, the shorter the “lifetime” of the state – the fuzzier the lines will be, and the less precisely its energy can be specified. Electronic transitions (e.g. from lower to higher energy) are quite fast (around 10 nanoseconds) and have relatively broad natural linewidths of about $5 \times 10^{-4}\text{ cm}^{-1}$, while a rotational state lifetime is about 10^3 seconds, corresponding to a relatively narrow natural linewidth of $5 \times 10^{-15}\text{ cm}^{-1}$. (See Atkins.) Measurement over infinity (all possible times) will give a precise energy value, but for any time period less than infinity the energy (hence the spectral lines) will be imprecise. Spectral linewidths will also vary depending on how “localized” the molecular constituents are. The more wavelike (less “localized”) a particle is, the less precisely its energy level can be specified.

that the molecule has. For example, different energy levels of electrons can be correlated with different orbitals, and these in turn are interpreted in terms of classical bonds. In a word, the spectra represent energy-level differences within the molecule, and the energy differences are interpreted in terms of classical structures.

No one, not even Woolley, doubts that the interpretations of spectra in terms of classical molecular structures are reliable. He does not reject the accuracy of the spectra as an indirect measurement of the interactions of particles within a molecule. What he rejects is the inference from these indirectly observed energy levels to a *classical* structure. In effect, he rejects the ontological assumptions of the classical model that go beyond the characterization of the molecule as a bundle of kinetic and potential energy.²⁴⁶

Premise 3. *These interpretations of our experimental results in terms of classical structures are not consistent with quantum mechanics.*

The inference from spectra to classical structure is inconsistent with quantum mechanics for the following reason. If molecules have classical structures, then the different energy levels that are interpreted as corresponding to classical structures would exist even if the molecule were isolated. As stated above in the description of the classical model, a classical structure is one that would exist even if the molecule were not being observed. But the reference to observation is inessential; it is often generalized to

²⁴⁶Woolley (1985: 1082) remarks: "One of the most important ideas to be clear about is that there is a definite philosophical content commonly associated with classical chemistry. Van't Hoff's stereochemistry is clearly founded on his profound belief in the reality of molecules as material objects in ordinary 3-dimensional ("physical") space; molecules could not be perceived directly simply because of the limitations inherent in our senses. For van't Hoff, stereochemistry was part of an argument to give a "proof" of the physical reality of molecules; the argument, however, is no more than a statement of an *a priori* philosophical position (classical realism). And if it is said that the "physical reality" of atoms and molecules is no longer disputed, this really only means that there is a more or less universal acceptance (conscious or unconscious) by physical scientists of the metaphysics of classical realism." Although he concludes in this article that structure is a "metaphor", and "not an object of belief", his remarks in (1978) are not anti-realist, just anti-classical.

include any exchange of energy with the environment.²⁴⁷ The important difference is not between unobserved and observed molecules, but between isolated and non-isolated molecules. Thus, if molecular structure is classical, molecules should exhibit the energy differences that correspond to structure even when they are in stationary (isolated) and not in transitional (non-isolated) states. Moreover, in spectroscopy, such energy differences *are* observed; identical nuclei are *not* fully symmetric or interchangeable.

Yet quantum mechanics tells us that identical nuclei in an isolated molecule cannot differ in their energy: two carbon nuclei cannot be said to occupy different fixed positions in the molecule, because this is equivalent to asserting that there exists a physically significant asymmetry between them. Now if molecules have classical structures, two carbon nuclei in different fixed positions *will* have different dynamical properties *even in isolation*. The symmetries that quantum mechanics predicts for isolated molecules will be broken if the classical model is correct.

So, Woolley claims, the fact that we *observe* energy differences in molecules does *not* justify the inference to the existence of a classical structure. It is one thing to observe differences in spectral lines and infer to different energy levels within a molecule. That in itself is no problem. It is another thing to think that these differences correspond to bonds between nuclei that occupy distinct, fixed relative positions within the molecule, an

²⁴⁷For example, Claverie and Diner (1980: 64) take pains to divorce environment-dependence from measurement-dependence: “[W]e content that, if the measurement process is really to play any role, it is only inasmuch as it entails some physical interaction between the measuring apparatus and the molecular system, which causes some modification of the state of the latter and it is not at all because the “measurement act” has some magic virtue of its own.” Ramsey (2000: 119) also stresses: “If a molecule were to experience conditions in nature like those it experienced in the measuring apparatus, the shape would also vary.”

interpretation of the data that implicitly asserts a claim about the nature of molecular reality.²⁴⁸ This is where Woolley gets off the boat, to the consternation of many:

[Woolley's analysis] clearly flies in the face of decades of experimental work. We have a number of techniques for the accurate determination of molecular bond lengths and angles, dipole moments, and optical activities, and have compiled impressive lists of these and other molecular parameters. How can quantum theory parade a string of structureless molecules before such an imposing jury of experimental facts? (Weininger 1984: 939).²⁴⁹

Premise 4. Experimental results in chemistry should be interpreted consistently with quantum mechanics.

The vast majority of chemists accept the basic correctness of quantum mechanics, so this premise is not really in dispute. But they also aware of decades of experimental work that have amply demonstrated the reliability of the classical interpretation of observations of molecules. So there is significant motivation to address the problems that Woolley and others have raised. The ubiquity of isomerism makes resolving the issues particularly pressing, for if identical particles are indistinguishable in quantum mechanics, then isolated isomers will be indistinguishable. Since isomers have very different causal powers, any adequate conception of molecular structure must enable chemistry to continue to draw the distinctions among isomers that it needs to.²⁵⁰

²⁴⁸ Woolley (1985: 1083) remarks: "[I]t is an elementary point of logic that seems easily overlooked that the success of this [classical] model in accounting for chemical observations *does not imply* that the molecular structure hypothesis is a *true* account of the nature of chemical substances as they "really" are (of "reality"); to go down that road is to become entangled in the mirages of metaphysics. The metaphysics of classical realism is optional. Molecular structure is a concept for solving chemical problems, not an object of belief."

²⁴⁹ Weininger ultimately agrees with Woolley's argument, however. He adds: "It is worth recalling the manner in which molecular structure parameters are obtained. Data such as bond lengths and bond angles are usually derived from either diffraction or spectroscopic experiments. Interpreting the output of these experiments, such as Patterson maps or rotational constants, requires that we *assume* the existence of discrete atoms in molecules. In other words, the classical picture of molecular structure is supported in part by data which require that picture for their existence."

²⁵⁰ Regarding the need for a viable structure concept, Claverie and Diner (1980: 58-59) remark: "The existence of optically active molecules, or more generally of stereoisomers, offers a striking example of this necessity. If all molecules were described by the ground state of their associated Hamiltonian with full

Premise 5. *So we should abandon the classical model as false.*

Quantum mechanics and classical chemistry give us conflicting answers as to what we should expect to observe in isolated molecules. Quantum mechanics should trump the classical picture, but experimental results appear to support the classical view. The conflict can be resolved, Woolley and others suggest, if we take note of the fact that our experiments involve non-isolated molecules while the quantum mechanical results assume an isolated molecule. That is, we can explain the *prima facie* disagreement between experiment and quantum mechanics if we claim that the energy differences that are not expected in isolated molecules appear as a result of the interactions with the environment that all observed molecules enter into. Structure appears when the molecule is perturbed by the environment, but will not appear when it is isolated.

This hypothesis allows us to explain our experimental results consistently with quantum mechanics. However, it follows that structure is not an intrinsic, context-invariant, property of a molecule, and the classical structure model should be abandoned:

[I]f one starts from a description of a molecule as an isolated, dynamical system consisting of the number of electrons and nuclei implied by the stoichiometric formula that interact via electromagnetic forces, one cannot even calculate the most important parameters in chemistry, namely, those that describe the molecular structure. This fact does not mean that “the quantum theory is wrong”, nor does it deprive the idea of molecular structure of its usefulness, for the key word in the above *ab initio* description is “isolated”: it does mean that it is wrong to regard molecular structure as an intrinsic property of a molecule. (Woolley 1978: 1074)²⁵¹

symmetry, no such isomers would exist as distinct species, since they would simply correspond to equally probable nuclear configurations of the same “complete” eigenstate.”

²⁵¹ Although the molecular structure debate continues, Woolley’s conclusion has stood up under peer assessment, and other arguments have also been given that yield the same end result.

Claverie and Diner (1980: 61) review Woolley’s work and conclude as follows: “[The] “classical” kind of structure is not easy to reconcile with the fully symmetric and time-independent picture suggested by the complete molecular eigenstate. It is this kind of structure that Woolley claims is not a permanent and intrinsic property of the molecule under any circumstances, particularly under conditions of high isolation from other molecules. If quantum mechanics is assumed to be fully valid, this claim would indeed seem to

If this is right, then efforts in chemistry to isolate molecules *to get greater resolution of their classical structure* are totally misconceived: instead, the more isolated molecule gets, or the less perturbation it undergoes, the more symmetric it will be.²⁵² It would also follow that molecules that have isomers, such as C_3H_4 , will not have in isolation the causal powers associated with each chemical species. Woolley notes that recent technical advances make it feasible to study some molecules in virtual isolation. So operational tests of the hypothesis are not far off.

be correct. Actually, it even becomes non-trivial to understand how the classical structure concept can nevertheless appear to be a valid one under suitable conditions."

Ramsey (1997, 2000) provides an independent argument to the same conclusion. He argues that the structure we observe depends on the time-frame of measurement: "Shape takes time to be observed." Our measuring instruments interact (exchange energy) with molecules over time, not instantaneously as in an ideal measurement. The amount of disturbance of the molecule's energy levels depends on the energy imparted to it by the measuring instrument and for how long the interaction occurs. It follows that measuring instruments that interact with the molecule for different lengths of time – that is, have different resolutions – will measure different energy levels. So they will have different structures, since structure corresponds to distinct energy levels in a molecule. For example, ammonia (NH_3) appears as a trigonal pyramid in IR spectroscopy and as two overlapping trigonal pyramids (hourglass-like) in a spectroscope that can distinguish rotation-vibration levels. If we had infinite time for measurement, we would observe fully symmetric molecules; this would correspond to the exact energy of an isolated molecule calculated by using the complete molecular Hamiltonian, as per Woolley's method. The ammonia molecule would look like a sphere at this limit of perfect spectral resolution, due to rotational symmetry. But at times less than infinity, the symmetry will be broken by our measuring instruments in different ways depending on time of interaction involved. Ramsey argues that measurement-relativity does not entail anti-realism about structure, but it does entail that structure is not intrinsic and not classical (2000: 119-121): "If a molecule has a shape, it has shape due to its interaction with other things in the world. Sometimes the interaction is with us and our instruments; many times it is not."

²⁵² Weininger (1984: 940) remarks: "Conventional wisdom tells us that by isolating a molecule and more finely resolving its energetic transitions we should be able to increase our detailed knowledge of its structure. However, if the analysis of Woolley and others is correct, then we may learn less about a molecule's structure by isolating it than we do under less extreme experimental conditions."

BIBLIOGRAPHY

- Adams, Frederick. (1986) "Intention and Intentional Action: The Simple View", *Mind & Language* vol. 1 no. 4 (winter), 281-301.
- Adams, Frederick. (1993) "Fodor's Modal Argument", *Philosophical Psychology* vol. 6 no. 1, 41-57.
- Ariew, Andre. (1996) "Innateness and Canalization", *Philosophy of Science* 63 supplement (Proceedings of the PSA, Part I: Contributed Papers), S19-S27.
- Armstrong, D.M. (1978) Nominalism and Realism: Universals and Scientific Realism Vol. I (New York: Cambridge University Press).
- Atkins, Peter W. (1998) Physical Chemistry, 6th ed. (New York: W. H. Freeman and Company).
- Baker, Lynne Rudder. (1995a) Explaining Attitudes (New York: Cambridge University Press).
- Baker, Lynne Rudder. (1995b) "Metaphysics and Mental Causation", in Heil and Mele, 75-96.
- Baker, Lynne R. (1997) "Why Constitution is Not Identity" *The Journal of Philosophy* vol. 94 no. 12 (December), 599-621.
- Baker, Lynne Rudder. (2000) Persons and Bodies: A Constitution View (New York: Cambridge University Press).
- Banich, Marie T. (1997) Neuropsychology: The Neural Bases of Mental Function (New York and Boston: Houghton Mifflin Company).
- Barron, L.D.. (1981) "Comment: Fundamental Symmetry Aspects of Optical Activity", *Chemical Physics Letters* vol. 79, no. 2 (April 15), 392-394.
- Bates, Elizabeth. (1994) "Modularity, Domain Specificity and the Development of Language", *Discussions in Neuroscience* vol. 10: 136-55. Reprinted in Bechtel et al. Philosophy and the Neurosciences (Oxford: Blackwell Publishers Inc., 2001).
- Bechtel, William. (1988) Philosophy of Mind: An Overview for Cognitive Science (Hillsdale, N.J.: Lawrence Erlbaum Associates).
- Bechtel, William and Jennifer Mundale (1999). "Multiple Realizability Revisited: Linking Cognitive and Neural States", *Philosophy of Science* 66 (June), 175-207.
- Bennett, Jonathan. (1971) Locke, Berkeley, Hume: Central Themes (Hong Kong: Oxford

University Press).

- Bennett, Jonathan. (1988) Events and Their Names (Indianapolis: Hackett Publishing Co.).
- Bhushan, Nalini and S. Rosenfeld, eds. (2000) Of Minds and Molecules: New Philosophical Perspectives on Chemistry (Oxford: Oxford University Press).
- Block, Ned. (1980) "What Is Functionalism?", in Block, N., ed., Readings in the Philosophy of Psychology Vol. 1 (Cambridge: Harvard University Press), 171-184.
- Block, Ned. (1986) "Advertisement for a Semantics for Psychology", in P. French, T. Uehling and H. Wettstein, Midwest Studies in Philosophy X: Studies in the Philosophy of Mind (Minneapolis: University of Minnesota Press), 615-678.
- Block, Ned. (1990) "Can the Mind Change the World?", in Boolos, G., ed. (1990) Meaning and Method: Essays in Honor of Hilary Putnam, (Cambridge: Cambridge University Press), 137-170.
- Block, Ned. (2003) "Do Causal Powers Drain Away?", *Philosophy and Phenomenological Research* 67, 1, (July 2003).
- Boyd, R. (1999) "Kinds, Complexity, and Multiple Realization" *Philosophical Studies* 95: 67-98.
- Brown, C. ed., (1990). Chirality in Drug Design and Synthesis (San Diego: Academic Press, Inc.)
- Burge, Tyler. (1977) "Belief *De Re*", *The Journal of Philosophy* vol. 74 no. 6 (June), 338-362.
- Burge, Tyler. (1979) "Individualism and The Mental," in French, Uehling, Jr. and Wettstein, eds. Midwest Studies in Philosophy IV: Studies in Metaphysics (Minneapolis: University of Minnesota Press).
- Burge, Tyler. (1982) "Other Bodies", in Woodfield, A., ed., Thought and Object (New York: Oxford University Press), 97-120.
- Burge, Tyler. (1986) "Individualism and Psychology", *The Philosophical Review* vol. 95 no.1 (January), 3-45.
- Burge, Tyler. (1989) "Individuation and Causation in Psychology", *Pacific Philosophical Quarterly* vol. 70, 303-322.
- Burge, Tyler. (1995) "Mind-Body Causation and Explanatory Practice," in Heil and

Mele, 97-120.

- Cartwright, Nancy. (1983) How the Laws of Physics Lie (New York: Oxford University Press).
- Chomsky, Noam. (1967) "Recent Contributions to the Theory of Innate Ideas", *Synthese* 17, 2-11. Reprinted in S. Stich, ed. Innate Ideas (Berkeley and Los Angeles: University of California Press, 1975). Page numbers refer to reprint.
- Clark, Andy and D. Chalmers. (1998) "The extended mind", *Analysis* 58.1 (January), 7-9.
- Claverie, Pierre and Simon Diner. (1980) "The Concept of Molecular Structure in Quantum Theory: Interpretation Problems", *Israel Journal of Chemistry* vol. 19, 54-81.
- Christiansen, David (1992) "Causal Powers and Conceptual Connections", *Analysis* vol. 52 no. 3 (July), 163-68.
- Churchland, Paul. (1981) "Eliminative Materialism and the Propositional Attitudes", *The Journal of Philosophy* vol. 78 no. 2 (February 1981), 67-90; reprinted in Rosenthal, ed., 601-612.
- Churchland, Paul. (2002) "Outer Space and Inner Space: The New Epistemology", Proceedings and Addresses of The American Philosophical Association vol. 76 no. 2 (November), 25-48.
- Crossley, Roger (1995). Chirality and the Biological Activity of Drugs (Boca Raton: CRC Press, Inc.
- Cummins, Robert (1988) "Comments on Dretske", in R. Grimm and D. Merrill, Contents of Thought (Tucson: University of Arizona Press), 44-55.
- Cummins, Robert and Denise D. Cummins, eds. (2000) Minds, Brains and Computers: The Foundations of Cognitive Science (Malden and Oxford: Blackwell Publishers Ltd.).
- Davidson, Donald. (1963) "Actions, Reasons and Causes", *The Journal of Philosophy* vol. 60; reprinted in Davidson (1980), 3-19.
- Davidson, Donald. (1971) "Agency", in Binkley, Bronaugh and Marras, eds., Agent, Action and Reason (Toronto: University of Toronto Press); reprinted in Davidson (1980), 43-61.
- Davidson, Donald. (1980) Essays on Action and Events (New York, Oxford University Press).

- Davidson, Donald. (1985) "Reply to Patrick Suppes" in Vermazen and Hintikka, 247-252.
- Devitt, Michael. (1996) Coming to Our Senses (New York: Cambridge University Press).
- Devitt, Michael. (2001) "Sustaining Actualism", *Philosophical Perspectives* 15, Metaphysics, 415-419.
- Devitt, M. and K. Sterelny. (1987) Language & Reality (Cambridge: The MIT Press).
- Dretske, Fred. (1988) Explaining Behavior (Cambridge: The MIT Press).
- Dretske, Fred. (1995) Naturalizing the Mind (Cambridge: The MIT Press).
- Dunn, J.M. (1990). "Relevant Predication 2: Intrinsic Properties and Internal Relations", *Philosophical Studies* vol. 60, 177-206.
- Egan, Frances. (1994) "Individualism and Vision Theory", *Analysis* vol. 54 no. 4 (October), 258-264.
- Egan, Frances. (2003) "Putting Mind in Its Place" (unpub. MS), presented at the CUNY Cognitive Science Symposium and Discussion Group, Nov. 14, 2003.
- Field, H. (1978) "Mental Representation", *Erkenntnis* 13 no. 1, 9-61; reprinted in Block, ed, (1980) Readings in the Philosophy of Psychology Vol. 1 (Cambridge: Harvard University Press), and in Field (2001).
- Field, Hartry. (1988) "'Narrow' Aspects of Intentionality and the Information-Theoretic Approach to Content". (Reprint)
- Field, Hartry. (1994) "Deflationist Views of Meaning and Content", *Mind* vol. 103 no. 411 (July), 249-285.
- Field, Hartry. (2001) Truth and the Absence of Fact (New York: Oxford University Press).
- Fodor, Jerry. (1975) The Language of Thought (Cambridge: Harvard University Press).
- Fodor, Jerry. (1980) "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology," *Brain and Behavioral Sciences* III, 1 (March 1980). Reprinted in Rosenthal, 485-498.
- Fodor, Jerry. (1987) Psychosemantics (Cambridge: The MIT Press).
- Fodor, Jerry. (1989) "Making Mind Matter More", *Philosophical Topics* vol. 17 no. 1

- (Spring), 59-79. Reprinted in A Theory of Content (Cambridge: The MIT Press, 1992), 136-159.
- Fodor, Jerry. (1990) A Theory of Content (Cambridge: The MIT Press).
- Fodor, Jerry. (1991) "A Modal Argument for Narrow Content" *The Journal of Philosophy* vol. 88 no. 1 (January), 5-26.
- Fodor, Jerry. (1994) The Elm and The Expert (Cambridge: The MIT Press).
- Fodor, Jerry. (2000) The Mind Doesn't Work That Way (Cambridge: The MIT Press).
- Fodor, J. and Zenon Pylyshyn. (1988) "Connectionism and Cognitive Architecture: A Critical Analysis", *Cognition* vol. 28, 3-71.
- Francescotti, Robert. (1999) "How to Define Intrinsic Properties", *Noûs* vol. 33 no. 4, 590-609.
- Geach, Peter. (1969) God and the Soul (London: Routledge and Kegan Paul).
- Godfrey-Smith, P. (1998). "Semantics, teleological", in E. Craig, ed., *Routledge Encyclopedia of Philosophy* (London: Routledge). Retrieved from <http://www.rep.routledge.com/article/W039>.
- Goldman, Alvin. (1970) A Theory of Human Action (Englewood Cliffs, N.J.: Prentice-Hall Inc.)
- Heil, John and Alfred Mele, eds. (1995) Mental Causation (New York: Oxford University Press).
- Hempel, Carl. (1948) Aspects of Scientific Explanation (Baltimore: Williams and Wilkins Co.).
- Herrick, Paul. (2000) The Many Worlds of Logic, 2nd ed. (Orlando: Harcourt Brace & Company).
- Hicks, Janice, ed. (2002) Chirality: Physical Chemistry (Washington, D.C.: American Chemical Society; distributed by Oxford University Press).
- Hitchcock, Christopher. (1993) "A Generalized Probabilistic Theory of Causal Relevance", *Philosophical Studies* vol. 78, 335-64.
- Horgan, Terence. (1997) "Kim on Mental Causation and Causal Exclusion", in J. Tomberlin, ed., Philosophical Perspectives, 11: Mind, Causation and World (Malden, Mass.: Blackwell Publishers, Inc.), 165-184.

- Horgan, Terence and Michael Tye (1985) "Against the Token Identity Theory" in E. Lepore and B. McLaughlin, eds. Actions and Events: Perspectives on the Philosophy of Donald Davidson (New York: Basil Blackwell Ltd.), 427-443.
- Horgan, Terence and James Woodward. (1991) "Folk psychology is here to stay", in Greenwood, J., ed., The Future of Folk Psychology (New York: Cambridge University Press), 149-175.
- Hornsby, Jennifer. (1980a) "Which Physical Events are Mental Events?" *Proceedings of the Aristotelian Society*, London (Dec. 18, 1980), 73-92.
- Hornsby, Jennifer. (1980b) Actions (London: Routledge and Kegan Paul Ltd.).
- Hornsby, Jennifer. (1986) "Physicalist Thinking and Conceptions of Behaviour", in Pettit, P. and J. McDowell, eds. Subject, Thought and Context (New York: Oxford University Press), 95-115.
- Hornsby, Jennifer. (1997) Simple Mindedness: In Defense of Naive Naturalism in the Philosophy of Mind (Cambridge: Harvard University Press.).
- Humberstone, I.L. (1996) "Intrinsic/Extrinsic" *Synthese* vol. 108, 205-267.
- Hurley, Susan (1998) "Vehicles, contents, conceptual structure and externalism", *Analysis* vol. 58 no. 1 (January), 1-6.
- Jackson, Frank and Philip Pettit. (1988) "Functionalism and Broad Content", in *Mind* vol. 97 no. 387 (July), 381-400.
- Jacques, Jean (1993). The Molecule and Its Double (New York: McGraw-Hill Inc.; trans. by Lee Scanlon, The Language Service, Inc., Poughkeepsie, N.Y.)
- Janoschek, Rudolf, ed. (1991). Chirality: From Weak Bosons to the α -Helix (Berlin and Heidelberg: Springer-Verlag).
- Johnston, Mark. (1992) "Constitution is Not Identity", *Mind* vol. 101 no. 401 (January 1992), 89-105.
- Kim, Jaegwon. (1982) "Psychophysical Supervenience", *Philosophical Studies* vol. 41, 51-70; reprinted in Kim (1993), 175-193.
- Kim, Jaegwon. (1984a) "Concepts of Supervenience", *Philosophy and Phenomenological Research* vol. 45, 315-26; reprinted in Kim (1993), 53-78.
- Kim, Jaegwon. (1984b) "Self-Understanding and Rationalizing Explanations", *Philosophia Naturalis* vol. 21, 309-320.

- Kim, Jaegwon. (1992) "Multiple Realization and the Metaphysics of Reduction", *Philosophy and Phenomenological Research* 52, 1-26; reprinted in Kim (1993), 309-335.
- Kim, Jaegwon. (1993) Supervenience and Mind: Selected Philosophical Essays (New York: Cambridge University Press).
- Kim, Jaegwon. (1998) Mind in a Physical World (Cambridge: The MIT Press).
- Lange, Marc. (2002) An Introduction to the Philosophy of Physics (Oxford: Blackwell Publishers Ltd.).
- Langton, Rae and David Lewis. (1998) "Defining 'Intrinsic'", *Philosophy and Phenomenological Research* vol. 58 no. 2 (June), 333-345.
- Latal, H. (1991) "Parity Violation in Atomic Physics", in Janoschek, ed., 1-17.
- Le Poidevin, Robin. (1994) "The Chemistry of Space", *Australasian Journal of Philosophy* vol. 72 no. 1 (March), 77-87.
- Le Poidevin, Robin. (2000) "Space and the Chiral Molecule", in Bhushan and Rosenfeld, ed., 129-142.
- LePore Ernest and Barry Loewer. (1987) "Mind Matters", *Journal of Philosophy* 84 (11) (November), 630-642.
- Lewis, David. (1983) "Extrinsic Properties", *Philosophical Studies* vol. 44, 197-200.
- Liboff, Richard. (2003) Introductory Quantum Mechanics, 4th ed. (San Francisco: Pearson Education, Inc. publishing as Addison Wesley.)
- Loar, Brian. (1988) "Social Content and Psychological Content", in R. Grimm and D. Merrill, eds., Contents of Thought (Tucson: University of Arizona Press), 99-110.
- Loewer, Barry. (2002) "A Guide to Naturalizing Semantics", unpublished. Presented at the New York University Spring 2002 Mind and Language Seminar.
- Lycan, William. (1988) Judgment and Justification (Cambridge: Cambridge University Press).
- MacDonald, Cynthia. (1989) Mind-Body Identity Theories (London and New York: Routledge).
- Mackie, John. (1965) "Causes and Conditions" *American Philosophical Quarterly* vol. 2/ 4 (October), 244-55 and 261-4. Reprinted in Sosa and Tooley.

- McDowell, John. (1985) "Functionalism and Anomalous Monism", in E. Lepore and B. McLaughlin, eds. Actions and Events: Perspectives on the Philosophy of Donald Davidson (New York: Basil Blackwell Ltd.), 387-398.
- McGinn, Colin. (1989) Mental Content (Oxford: Basil Blackwell Ltd.).
- McLaughlin, Brian. (1989) "Type Epiphenomenalism, Type Dualism, and the Causal Priority of the Physical", in J. Tomberlin, ed., Philosophical Perspectives, 3: Philosophy of Mind and Action Theory (Atascadero, Calif.: Ridgeview Publishing Co.), 109-135.
- McLaughlin, B., ed. (1991) Critical Essays on the Philosophy of Fred Dretske (Oxford: Basil Blackwell).
- McLaughlin, Brian. (1995) "On Davidson's Response to the Charge of Epiphenomenalism" in Heil and Mele, 27-40.
- Millikan, Ruth. (1986) "Thoughts Without Laws; Cognitive Science Without Content" *The Philosophical Review* vol. 95 no. 1 (January), 47-80.
- Millikan, Ruth. (1993) White Queen Psychology and Other Essays for Alice (Cambridge: The MIT Press).
- Millikan, Ruth. (1995) "Explanation in Biopsychology", in Heil and Mele, 211-232.
- Nerlich, Graham. (1994) The Shape of Space, 2nd ed. (Cambridge: Cambridge University Press).
- Patsakos, George. (1976) "Classical Particles in Quantum Mechanics", *American Journal of Physics* vol. 44 no. 2 (February), 158-166.
- Paul, L.A. (2000) "Aspect Causation" *The Journal of Philosophy* vol. 97 no. 4 (April), 235-256.
- Peacocke, Christopher. (1993) "Externalist Explanation", *Proceedings of the Aristotelian Society New Series* vol. 93 (Bristol: The Aristotelian Society), 203-230.
- Pereboom, Derk. (1991) "Why a scientific realist cannot be a functionalist", *Synthese* vol. 88, 341-358.
- Pereboom, Derk. (2002) "Robust Nonreductive Materialism", *The Journal of Philosophy* vol. 99 no. 10 (October), 499-531.
- Pereboom, Derk and Hilary Kornblith. (1991) "The Metaphysics of Irreducibility", *Philosophical Studies* vol. 63 no. 2 (August), 125-145.

- Pettit, Philip. (1986) "Broad-Minded Explanation and Psychology", in Pettit and McDowell, eds., 17-58.
- Pettit, Philip. and John McDowell, eds. (1986) Subject, Thought and Context (New York: Oxford University Press).
- Plantinga, Alvin. (1973/1979) "Transworld Identity or Worldbound Individuals?", in M. Munitz, ed., Logic and Ontology (New York: New York University Press); reprinted in M. Loux, ed., The Possible and the Actual (Ithaca: Cornell University Press, 1979), 146-165.
- Primas, H. (1980) "Foundations of Theoretical Chemistry", in Woolley, R.G., ed., Quantum Dynamics of Molecules: The New Experimental Challenge to Theorists (New York: Plenum Press); cited in Weininger (1984).
- Putnam, Hilary. (1975a) "The Meaning of 'Meaning'", reprinted in Mind, Language and Reality: Philosophical Papers vol. II (New York: Cambridge University Press), 215-271.
- Putnam, Hilary. (1975b) "Philosophy and our mental life", in Mind, Language and Reality: Philosophical Papers vol. II (New York: Cambridge University Press), 291-303.
- Putnam, Hilary. (1975c) "Brains and Behavior" in Mind, Language and Reality, Philosophical Papers vol. II (New York: Cambridge University Press), 429-40.
- Pylyshyn, Z. (1984) Computation and Cognition (Cambridge: The MIT Press).
- Quine, W.V.O. (1960) Word & Object (Cambridge: The MIT Press).
- Ramsey, Jeffry. (1997) "Molecular Shape, Reduction, Explanation and Approximate Concepts", *Synthese* 111, 233-251.
- Ramsey, Jeffry. (2000) "Realism, Essentialism, and Intrinsic Properties: The Case of Molecular Shape", in Bhushan and Rosenfeld, eds., 117-128.
- Rea, Michael. (1995) "The Problem of Material Constitution", *The Philosophical Review* vol. 104 no. 4 (October), 525-552.
- Rey, Georges. (1997) Contemporary Philosophy of Mind (Cambridge: Blackwell Publishers Inc.).
- Robinson, William. (1990) "States and Beliefs", *Mind* vol. 99 no. 393 (January), 33-51.
- Rosenberg, Alexander. (1974) "On Kim's Account of Events and Event-Identity",

Journal of Philosophy vol. 71 no. 11 (June 13), 327-36.

- Rosenthal, David, ed. (1991a) The Nature of Mind (New York: Oxford University Press).
- Rosenthal, David. (1991b) "The independence of consciousness and sensory quality", in Villanueva, E., ed. Consciousness (Atascadero: Ridgeview), 15-36.
- Ross, Patricia. (1999) "The Limits of Physicalism", *Philosophy of Science* vol. 66 (March), 94-116.
- Salmon, M., J. Earman, C. Glymour, J. Lennox, P. Machamer, J. McGuire, J. Norton, W. Salmon and K. Schaffner. (1992/1999) An Introduction to the Philosophy of Science (Indianapolis: Hackett Publishing Co.; first published by Prentice-Hall Inc.).
- Shapiro, Lawrence. (2000) "Multiple Realizations", *The Journal of Philosophy* vol. 97 no. 12 (December), 635-54.
- Schiffer, Stephen. (1981) "Truth and the Theory of Content", in Parret, H. and J. Bouveresse, eds., Meaning and Understanding (Berlin: Walter de Gruyter), 204-222.
- Schiffer, Stephen. (1987) Remnants of Meaning (Cambridge: The MIT Press).
- Schiffer, Stephen. (1991) "Ceteris Paribus Laws", *Mind* vol. 100 no. 397 (January 1991), 1-17.
- Segal, Gabriel. (1989) "Seeing What Is Not There," *The Philosophical Review* vol. 98 no. 2 (April), 189-214.
- Sosa, Ernest and M. Tooley, eds. (1993) Causation (New York: Oxford University Press).
- Steward, Helen. (1997) The Ontology of Mind (New York: Oxford University Press).
- Stich, Stephen. (1975) "Introduction: The Idea of Innateness", in Stich, S., ed., Innate Ideas (Berkeley and Los Angeles: University of California Press), 1-22.
- Stich, Stephen. (1978) "Autonomous Psychology and the Belief-Desire Thesis", *The Monist* 61 no. 4 (October 1978). Reprinted in Rosenthal, 590-600.
- Stich, Stephen. (1983) From Folk Psychology to Cognitive Science (Cambridge: The MIT Press).
- Stich, Stephen. (1991) "Narrow Content Meets Fat Syntax," in Loewer, Barry and Georges Rey, eds. Meaning in Mind: Fodor and His Critics (Cambridge: Basil

- Blackwell), 239-254.
- Sosa, Ernest. (1984) "Mind-body Interaction and Supervenient Causation" in French, Uehling and Wettstein, eds., Midwest Studies in Philosophy IX: Causation and Causal Theories (Minneapolis: University of Minnesota Press), 271-281.
- Strawson, P.F. (1985) "Causation and Explanation", in Vermazen and Hintikka, 15-135.
- Suppes, Patrick. (1985). "Davidson's Views on Psychology as a Science", in Vermazen and Hintikka, 183-194.
- Thalberg, Irving. (1977) Perception, Emotion and Action: A Component Approach (New Haven: Yale University Press).
- Van Gelder, Tim. (1995). "What Might Cognition Be, If Not Computation?" *The Journal of Philosophy* vol. 92 no. 7 (July), 345-381.
- Van Gulick, Robert. (1995) "Who's In Charge Here? And Who's Doing All the Work?", in Heil, John and Alfred Mele, eds. Mental Causation (New York: Oxford University Press), 233-56.
- Vermazen, Bruce and Merrill Hintikka, eds. (1985) Essays on Davidson: Action and Events (New York: Oxford University Press).
- Waddington, C.H. (1975) The Evolution of an Evolutionist (Ithaca: Cornell University Press).
- Weininger, Steven. (1984) "The Molecular Structure Conundrum: Can Classical Chemistry be Reduced to Quantum Chemistry?" *Journal of Chemical Education* 61, 939-944.
- Weininger, Steven. (2000) "Butlerov's Vision: The Timeless, the Transient, and the Representation of Chemical Structure", in Bhushan and Rosenfeld, eds., 143-161.
- Woolley, R. (1978) "Must a Molecule Have a Shape?", *Journal of the American Chemical Society* 100, 1073-1078.
- Woolley, R. (1985) "The Molecular Structure Conundrum" *Journal of Chemical Education* 62, 082-1084.
- Zumdahl, Steven. (1996) Basic Chemistry, 3rd ed. (Lexington, MA: D.C. Heath and Company).