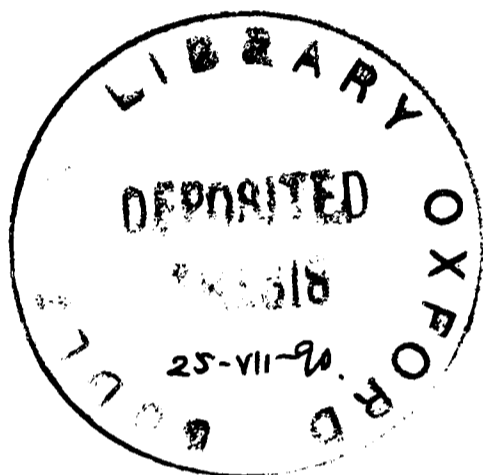


Anomalism, Supervenience, and
Explanation in Cognitive Psychology



Mark Rowlands
Jesus College
Oxford.

A Thesis Submitted to the University of Oxford in Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

Trinity Term, 1988

Anomalism, Supervenience, and Explanation in Cognitive Psychology

by

Mark Rowlands

of

Jesus College

A Thesis Submitted to the University of Oxford in Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

Abstract

This thesis defends the claim that the principle of *methodological solipsism* can play no role in the formation of the theories of cognitive psychology. Corresponding to this negative claim, but assuming a comparatively minor role, will be the positive claim that a scientific psychology ought to deal in explanations which relate mental states in virtue of their semantic contents.

The basis of the case against methodological solipsism is the claim that the explanatory properties invoked by this principle are individuation dependent on properties of semantic content. In Chapter I the idea of methodological solipsism will be discussed, and two forms distinguished. One of the versions of methodological solipsism identified invokes the explanatory notion of the *narrow content* of a mental state. The other version invokes the notion of formal or syntactic properties possessed by mental states. In both cases it will be argued that these properties can be identified only by way of the semantic contents of their associated mental states. The notion of narrow content will be discussed in Chapter II. The case against that version of methodological solipsism which invokes syntactic properties will be constructed in Chapters III-V. The latter argument constitutes the bulk of the thesis, and derives from considerations centred around the principles of anomalism, supervenience, and the relation between them. These arguments are intended to be of independent interest as solutions to certain persistent problems in the philosophy of mind.

ACKNOWLEDGEMENTS

I would like to express sincere gratitude to my supervisor Colin McGinn. Without Colin's extraordinary familiarity with, and deep appreciation of, the problems in this area, I would certainly not have been able to submit this thesis in the time in which it has been submitted. In my time with him, Colin has been a conscientious supervisor whose identification of problems, inconsistencies or tensions in the work has saved me months, if not years, of work.

I would also like to thank Dr. Galen Strawson whose comments, particularly on Chapter I, have made this thesis better than it would otherwise have been.

CONTENTS

	Introduction.	1
I	Causation, Causal Explanation, and Methodological Solipsism in Psychology.	4
	<i>Appendix I: Dennett's Illata/Abstracta Distinction as a Case of the Causal Power/Causal Relevance Distinction.</i>	54
II	Conceptual Explanatory Methodological Solipsism: Narrow Content, Expressibility, and Explanation.	66
III	The Anomalism and Supervenience of Mental Properties. Part 1: The Anomalism of Mental Properties. Part 2: The Supervenience of Mental Properties. A. Motivation. B. Characterization. Part 3: Anomalism and Supervenience: Compatibility.	93
IV	Anomalism, Supervenience, and The Dual Component Theory of Mental Properties. Part 1: A Possible Solution: Rationality and Realization. Part 2: Narrow Mental States and Rationality.	170
	<i>Appendix II: Reply to a Possible Objection: Ontological Versus Individuation Dependence.</i>	208
	<i>Appendix III: The Nature of Psychological Explanation: Normativism Versus Generalism.</i>	218
V	Formal Explanatory Methodological Solipsism: The role of Syntax in Psychological Explanation.	235
	Conclusion	280
	Bibliography	287

Introduction.

What form should theories of a scientific psychology take? The question identifies the topic with which this thesis is concerned. It is a philosophical rather than a psychological question. The question is concerned not with the content of any specific psychological theory, but with the form which all theories of a scientific psychology must, or should, take. And an answer to the question would not consist in any particular psychological theory, but stand on a meta-level relative to such theories. It would be a theory *about* psychological theories.

Answers to this question can be loosely divided into two opposing camps. On the one hand there are those who think that theories of a scientific psychology will be substantially similar to theories of ordinary folk psychology. They may differ in their range and level of precision, but they will be essentially similar sorts of theory. They will either invoke the same explanatory entities (beliefs, desires, semantic contents, etc.), or else invoke entities of fundamentally similar types. Scientific psychology, according to this view, is an extension and precisification of folk psychology. On the other hand, there are those who think that the theories of scientific and folk psychology will be radically different in kind. Proponents of this view purport to see something in folk psychological constructs, and the semantic contents upon which they are based, which makes them unsuitable for employment in a respectable scientific psychology. And the most

commonly identified problem in this regard is that folk psychological states fail to respect the principle of *Methodological Solipsism*.

The principle of methodological solipsism is a methodological constraint on the formation of theories of cognitive psychology. It has received many and diverse formulations in recent literature, but all are based around the idea that, for the purposes of a scientific psychology, mental states should be individuated purely by reference to what is 'in the head' of the organisms which are the subject of the psychology. Explication of the notion of being 'in the head' will take place in Chapter I. There it will be argued that the principle of methodological solipsism is tied to the notion of causal determination and explanation. Two importantly different forms of methodological solipsism will be distinguished. These will be known as *ontological* and *explanatory* methodological solipsism. It will be seen that only the explanatory version supplies any substantive methodological constraint. A further distinction will be drawn between what will be known as *formal* and *conceptual* explanatory methodological solipsism. This distinction turns on the sort of explanatory properties invoked by each. The remainder of the thesis will argue that neither version of explanatory methodological solipsism is acceptable.

Chapter II will discuss conceptual explanatory methodological solipsism. The explanatory properties invoked by this principle consist in what has become known as *narrow semantic content*. It will be argued that the notion of narrow content is incoherent, and that any narrow semantic properties possessed by a mental state must be necessarily

individuation dependent on ordinary semantic contents. Therefore the principle of conceptual explanatory methodological solipsism will be rejected.

Chapter III begins the long argument against the formal version of explanatory methodological solipsism. The goal of this argument will, again, be to show that the explanatory properties invoked by this principle - formal or syntactic properties - are necessarily individuation dependent on ordinary semantic contents. The basis of this argument will consist in considerations derived from a discussion of the relationship between the principles of anomalism and supervenience. The argument will be rehearsed in terms of the relation between mental and physical properties. These arguments are, at the same time, meant to be of independent interest as solutions to certain persistent problems in the philosophy of mind.

Subtleties aside, the conclusion of these arguments will consist in a principle of the following form: Suppose there exist two properties A and B. Then, if A is (i) anomalous, and (ii) supervenient on B, then the property B can only be individuated in terms of property A. Chapters III and IV will be concerned with establishing this principle in relation to mental and physical properties. Chapter V will apply the principle to the principle of formal explanatory methodological solipsism. The principle will be used to show that formal or syntactic properties, *at least as those properties are understood in the philosophy of psychology* are individuation dependent on semantic contents, or on the sentences we use to ascribe those contents.

Chapter I:

Causation, Causal Explanation, and Methodological Solipsism in Psychology.

1

This chapter will argue that in talking of causation and explanation, four principles must be distinguished. Part of the chapter will consist in a statement of the principles involved and a defence of the distinctions implied. The remainder will be concerned with the application of these principles, firstly to mental properties, and subsequently to the principle of methodological solipsism in psychology. The fourfold distinction is between the following:

1. Causal Efficacy. Causal efficacy is a first-order property possessed by event-tokens. It is in virtue of having the property of causal efficacy that one event-token can be a cause (or effect) of another.
2. Causal Power. Causal Power is a second-order property which can be had by first-order properties. A given (first-order) property has causal power if and only if it bestows on the event-tokens which possess it dispositions to enter into token causal relations. That is, the event-tokens which instantiate the property in question enter into these causal relations because that property has the appropriate causal power. Put in terms of definition (1): a given first-order property has causal

power if and only if bestows upon its instances the (first-order) property of causal efficacy.

3. Causal Relevance. Causal Relevance is the property of playing an explanatory role in a causal explanation. The primary bearers of causal relevance are other properties. A property can often have causal relevance in virtue of having an appropriate causal power. However, it will be argued that the possession of causal power by a property is neither a necessary nor, given certain assumptions, a sufficient condition, for that property's possessing causal relevance.

4. Explanatory Relevance. Explanatory relevance is the property of playing an explanatory role in an explanation of any sort, causal or otherwise. Thus explanatory relevance stands to causal relevance in a relation of genus-species: any property which possesses causal relevance thereby possesses explanatory relevance, but not vice versa. So causal and explanatory relevance are to be distinguished, assuming that not all explanation is causal. The primary bearers of explanatory relevance are again other properties.

The first point to note is that discussion of these and related matters will be framed within a property-exemplification account of events.¹ According to this account, an event-type is a property the

1. The property-exemplification account of events features largely in the work of Jaegwon Kim. See: "On the Psychophysical Identity Theory", in American Philosophical Quarterly, vol. 3, (1966), pp. 227-35. "Causation, Nomic Subsumption, and the Concept of Event" in The Journal of Philosophy, 70 (1973), pp. 27-36. "Events as Property Exemplifications", in M. Brand and D. Walton eds. Action Theory, (Dordrecht, Netherlands), pp. 159-77. Another property-exemplification account is provided by Lawrence Lombard: "Events", Canadian Journal of

exemplification of which is an event-token. Subtleties aside, an event-token is the exemplification of a property (or n-adic property or relation) by an object (or n-tuple of objects) at a time (or throughout an interval of time). On this view, events are dated concrete particular occurrences which can be represented by a description of the form $[x,P,t]$, where x is an object, t is a time of occurrence, and P is a property instantiated by x .² The type-token distinction adhered to throughout this thesis follows the property-exemplification account. Event-types will be regarded as properties, event-tokens will be regarded as instances of these properties. The term 'event' as used here will refer purely to event-tokens. Therefore, in this thesis, events will be understood as instances of properties.³

The somewhat non-standard interpretation of the notion of the *order* of a property again stems from the property-exemplification account. A given event-type P may be regarded (though not in every case) as a

Philosophy, vol. 9 (1979), pp. 425-60. Also Events: A Metaphysical Study. (London, Routledge and Kegan Paul, 1986).

2. The real position, as delineated by Kim, is of course, slightly more complicated. Events, according to Kim, can be given various non-equivalent descriptions. They can, however, also be described canonically. And it is the canonical description of an event which Kim characterizes by the description operator '[]'. Hence a description such a $[x,P,t]$ is a canonical description of an event; x is a constitutive object, P is a constitutive property, t is a constitutive time of the event.

3. Again, on Kim's account, there is a distinction between an event instantiating a property, and an event being an instance of a property. An event can instantiate any number of properties, but it can be an instance of one and only one property, its constitutive property, the property referred to in a canonical description of that event.

first-order property. In that case the property which is the being an exemplification of P in an object at a time is a second-order property. And the notion of *orders* of properties will be understood along these lines. Generalizing, a second-order property, certainly as it appears in the above definitions, is a *property of a property*. And a third-order property is, correspondingly, a property of a property of a property, etc.

This being understood, the definition of causal *efficacy* is just a reflection of the view that causation is a relation which holds between concrete particulars - which holds between events, here conceived of as instances of properties. This assumption, which has not gone entirely unchallenged, features prominently in the work of Davidson (minus the property-exemplification account of events, of course), and will be adhered to throughout this paper.⁴

The distinction between causal *efficacy* and causal *power* is just the type-token distinction applied to causation. That is, as understood here, the bearers of causal power are properties, not their instances. Given that causation is a relation which holds between events, it is still possible, indeed desirable, to allow that events enter into causal relations in virtue of certain properties they possess. Some have

4. See Donald Davidson; "Actions, Reasons, and Causes", The Journal of Philosophy, vol. 60 (1963), pp. 685-700. Reprinted in Davidson; Essays on Actions and Events (Oxford, Oxford University Press, 1980), pp. 3-19. "Causal Relations", The Journal of Philosophy, vol. 64 (1967), pp. 691-703. Reprinted in Davidson (1980), op. cit., pp. 149-62

called these properties the causally relevant properties of the events.⁵ I reserve the term 'causal relevance' for a different use. Instead, I will refer to them as *causally powerful* properties. Some (perhaps all) causally powerful properties are also causally relevant. But it will be argued that causal power and causal relevance are distinct properties. The intuitive model of causation underlying the notion of causal power is the following: an event *c* causes an event *e* in virtue of a property *C* of *c* and a property *E* of *e*. It is the events *c* and *e* which stand in the relation of causation. This is attested to by certain well known facts, such as the irrelevance of the descriptions by which events are picked out to the causal relations in which the events stand.⁶ The properties *C* and *E* stand in some connected, but importantly distinct, relation. One way of putting this would be to say that events enter into token causal relations, properties enter into type causal relations. The token causal relation is what I have called the relation of causal efficacy, the type causal relation is what I have called the relation of causal power.

The efficacy-power distinction, then, is meant simply to be an expression of the type-token distinction. In a future chapter it will be argued that this distinction has, in many recent discussions, been

5. See, for example, Ted Honderich; "The Argument for Anomalous Monism", *Analysis*, vol. 42, no. 1, (1982), pp. 59-64. Honderich's views will be discussed in Chapter III.

6. See Davidson: "Actions, Reasons, and Causes", *op. cit.*; "Causal Relations", *op. cit.* Also "Mental Events", in L. Foster and J.W. Swanson eds., *Experience and Theory* (University of Massachusetts Press and Duckworth, London, 1970) pp. 79-101.

consistently confused or ignored. This has led to considerable confusion.⁷

Given the assumption that not all explanation is causal explanation, it is difficult to find fault with the distinction between causal and explanatory relevance. This assumption does not find univocal support, but will be adhered to throughout this thesis. I have said that the primary bearers of causal/explanatory relevance are properties. I wish to remain non-committal on the question of whether they are the *exclusive* bearers. In all future talk, properties will be regarded as the bearers of causal/explanatory relevance. Nothing much will turn on the question of whether events can also possess causal/explanatory relevance independently of the properties they instance.

The distinction which needs the most defence, both because of its relative novelty, and because of the work it will do later in this thesis, is the distinction between causal power and causal relevance.

Demonstration of the distinctness of the properties of causal power and causal relevance might proceed in one of two ways. On the one hand, one could try to show that the possession of causal power by a property is not *necessary* for the possession of causal relevance by that property. On the other hand one could try to show that possession of the former is not *sufficient* for possession of the latter.

7. See Chapter III, Part 2A.

Consider, first, whether possession of causal power by a property is sufficient for possession of causal relevance by that property. Under certain assumptions, to be stated shortly, it can be seen that a property's possession of causal power might not be sufficient for that property's possession of causal relevance. An argument to this conclusion might run as follows:

Explanations are always relative to the level of knowledge of explainers: an explanation can be regarded as genuine only if it has appropriate connections with what the explainers already know. Therefore, there may be causal powers which we, the explainers, have not the conceptual resources to represent, characterize, or categorize. In this case, a property which possessed such a causal power would not, in virtue of that possession, possess causal relevance.

There are two possible, but mutually incompatible, responses to this argument. The attractiveness of each will depend on the extent to which one is prepared to be a realist with regard to causal powers. If one adopts an anti-realist stance, along the lines of the traditional construal of Hume,⁸ one would not find the argument convincing. The reason is that the argument presupposes *realism* with respect to causal powers. It presupposes that there could be causal powers which we are

8. That this is the traditional construal of Hume does not imply that it is the only construal, or the correct one. E.J. Craig, for example, argues persuasively that Hume was a realist with regard to causal powers. See, The Mind of God and the Works of Man, (Oxford, Oxford University Press, 1987) Ch.2.

unable to represent, characterize, or categorize. The second response, to some extent at least, presupposes realism with respect to causal powers, and proceeds by way of the distinction between *available* and *total* explanation. In particular, it might be argued that the lack of sufficiency of causal power for causal relevance exhibited in arguments of the above sort is symptomatic only of the fact that the explanations involved are not total explanations: they are not explanations of the (or some) ultimate ideal, complete, and objective theory, but merely explanations of theories whose explanatory concerns are restricted by our finite perspectives. Both these responses lead into difficult areas, and an adequate discussion would require more space than is available. Consequently, I will concentrate on the second route: the distinctness of causal power and causal relevance will be demonstrated by showing that the possession of causal power by a property is not *necessary* for that property's possessing causal relevance.

It seems fairly uncontroversial that possession by a property of causal power is not *necessary* for possession by that property of causal relevance. Examples of properties which possess causal relevance without causal power abound in both folk psychological and folk macrophysical explanations. Consider, for example, the use of *relational* properties in such explanations. Consider the property of belonging to Fred. Why does Sally want X? Because X belongs to Fred. This, suitably filled in with reports of Sally's other psychological states, might be regarded as a perfectly adequate causal explanation of folk psychology, despite the fact that the property of belonging to Fred cannot plausibly be viewed as a possessor of causal power. Similarly,

in folk macrophysics. Why did body A accelerate at the same rate as body B? Because they were acted upon by forces of equal magnitude. This might be regarded as a perfectly satisfactory causal explanation of folk macrophysics, even though the property of being a force of equal magnitude to that acting on A (or B) cannot be plausibly viewed as a possessor of causal power. Sometimes in ordinary folk explanations it is the absence of a property which is regarded as an explanatory factor. Why did Sam like the pizza? Because there were no anchovies on it. Why did the bridge collapse? Because there were no stanchion rods underneath it. Negative properties cannot be bearers of causal power.

There is a line of response to these commonplace observations which seeks to minimize their importance by construing the use of the highly relational and negative properties described as merely symptomatic of inherently flawed folk psychological/macrophysical explanation schemata. The idea is that such properties would be excluded from proper scientific explanations, presumably by some general methodological principles of the scientific enterprise, such as Occam's Razor. However, I shall argue against this claim. Even in the domain of what we now regard as respectable science, the explanatory properties invoked do not, generally, possess causal power.

The argument to this conclusion derives from two main sources. Both involve properties which play a part in the various sciences. These are:

(i) *Numbers.* Numbers, or, rather, properties of number have causal relevance even though they do not possess causal power.

(ii) *Scientific Explanatory Properties.* It will be argued that many of the properties invoked by science - temperature, energy, velocity etc. - have causal relevance but do not have causal power.

1. Numbers. A *property of number*, as understood here, is a property such as *being the number twenty* etc. It seems to be fairly uncontroversial that properties of number can figure in causal explanations, and, thereby, possess causal relevance. For example, imagine a detector system which is set up to register - by way of a red light, for example - those objects, and only those objects, which weigh exactly twenty kg. Then, given various background assumptions, we could the following sentence as expressing a genuine causal explanation:

" In machine M the red light appeared because there was placed upon it an object O which weighed exactly 20 kg. "

If the above sentence does express a genuine causal explanation, then properties of number possess causal relevance; for the property of weighing 20 kg plays an explanatory role in the sentence. However, intuitively, one would not want to claim that properties of number also have causal power. Numbers are non-physical abstract properties; they do not appear to be the sort of thing that can have causal power. So, intuitively, properties of number appear to be the sort of thing which can have causal relevance, but not the sort of thing that can have causal power.

Intuitions are notoriously fallible. However, in this case they can be supported by an account of how properties of number can have causal relevance in the absence of causal power. The account is based on the theory of measurement.⁹ Consider, first, how real numbers can be used to serve as a scale for the measurement of, for example, mass. The initial idea is that there exist certain properties and relations obtaining between massive objects. Call these *mass-relations*. These mass-relations can, in principle, be specified without reference to numbers. Then a representation theorem linking these (non-numerically specified) mass-relations with the system of real numbers is set up. The basis of this representation theorem lies in the notion of isomorphism. The idea, baldly put, is this: numerical relations holding between real numbers are isomorphic with intrinsic mass-relations holding between objects. Consequently, assigning real numbers to objects can be used as a convenient way of discussing intrinsic mass-relations holding between objects. I will say here that numbers index intrinsic properties of mass, and, in virtue of this indexing, talk of numbers can 'go proxy' for talk of intrinsic properties of mass.¹⁰ Similarly, talk of relations between numbers can be used as a convenient way of discussing mass-relations between objects.

9. This account is essentially the one given in the appendix to Hartry Field's "Mental Representation", Erkenntnis, vol. 13, pp. 9-61. Reprinted in Ned Block, ed., Readings in the Philosophy of Psychology vol.2, (Massachusetts, Harvard University Press), pp. 78-114. See also R. Stalnaker, Inquiry, (Massachusetts, MIT Press, 1984), Ch.1.

10. The expression 'indexing' is due to Brian Loar, Mind and Meaning, (Cambridge, Cambridge University Press, 1981). Loar's account will be discussed in Chapter II.

An important point to note is that the intrinsic mass-relations do not require the existence of real numbers. Moreover, the numbers are *extrinsic* to the objects they index. There is no claim that objects or mass-relations are intrinsically mathematical. What mathematicity they possess is the result of the imposition of an external mathematical framework upon them.¹¹ A corollary of this is that numbers are always essentially eliminable from scientific explanations. They have the status of heuristic devices, in principle eliminable in favour of the intrinsic properties and relations which they index.¹²

This account indicates how properties of number can possess causal relevance in the absence of causal power. The reason is that numbers can act as external indices of intrinsic mass-properties, and numerical relations can act as external indices of mass-relations. And these latter properties and relations are those which possess causal power. Properties of number can possess causal relevance, in the absence of causal power, because numbers index states which do have causal power. The notion of indexing will be examined more fully later,

11. This point has not been properly understood by Field. In the appendix to "Mental Representation", op. cit., Field speculates that the relation between propositions and internal representations might be analogous to the relation between numbers and intrinsic properties of objects. He claims that if this is so, the internal representations would inherit the structure of propositions - they would have logical structure. I think this argument fails to appreciate the *externality* of the indexing relation.

12. It will be seen that this in principle eliminability is shared by the other examples of causally relevant but non-causally powerful properties discussed. These are, on the one hand, what I have called *scientific explanatory properties* and, on the other, what Dennett has called *abstracta*. Discussion of Dennett's notion of abstracta will take place in Appendix I.

where it will be applied to mental properties and the sentences we use to ascribe them.

2. Scientific Explanatory Properties The discussion of what are here called *scientific explanatory properties* will draw heavily on the (as yet unpublished) work of Simon Blackburn. In a recent paper, Blackburn seeks to display certain common misconceptions concerning the nature of physical states or properties.¹³ This account, it will be argued, can be used to support the distinction between causal power and causal relevance.

The core of Blackburn's account is his opposition to what he labels the 'Tractarian View' of physical properties. This is "... the mistake of supposing that for any physical property there should be a story, in terms of the configuration of some constituent things, saying what it is."¹⁴ Blackburn argues that many physical properties - certainly many properties which play an explanatory role in physical theories - are not *constitutional* properties; they are not individuated by way of any supposed constituent physical structure. A few examples will help make this clear.

13. Simon Blackburn: "Losing Your Mind: Physics, Identity and Elimination" (Unpublished). The following discussion is heavily indebted to this paper.

14. "Losing Your Mind: Physics, Identity and Elimination", op. cit., p.10.

(a) *Temperature*. Blackburn discusses the example of temperature at some length. The discussion begins with Patricia Churchland's observation that the equation of temperature with mean kinetic energy of molecules does not generally hold. The identification does not apply to the temperature of solids, plasmas, or vacua. Churchland concludes from this that reduction can be domain specific: temperature can be one thing in a gas, another thing in solids, another thing in plasmas, and yet another thing in vacua.¹⁵

The problem with this type of response is, as Blackburn points out, that it leaves it very unclear why we are justified in talking of temperature, as a unitary phenomenon, at all. That is, it leaves it very unclear why we would be justified in attributing the predicate '... is of temperature ...' to solids, plasmas, gases etc. The Churchland response seems to leave us no basis for attributing the same predicate to different phase states.¹⁶

15. "Losing Your Mind: Physics, Identity and Elimination", op. cit., pp. 7-9. Churchland's original discussion is in P.S. Churchland, Neurophilosophy.

16. This point is made by Blackburn, "Losing Your Mind: Physics, Identity and Elimination", op. cit., p. 7. The point is familiar from discussions centering around functionalism and the variable realization of mental states. If human pain, for example, is a different brain state from Martian pain, then, so the argument goes, we cannot give any account either of what property it is in virtue of which humans and Martians can both be in pain, or of what property is common to all token states in virtue of which they are pain states. This point is well brought out by Michael Tye; "Functionalism and Type Physicalism", Philosophical Studies vol. 44 (1983), pp. 161-74. See especially p. 162.

Blackburn diagnoses the Churchland response as a case of the Tractarian View of properties. In contrast, Blackburn argues that a non-Tractarian view of temperature is implicated in the use of the concept in physical science. Physics can explain how solids, liquids, gases, plasmas, and vacua can be at the same temperature, but it does not do this by citing any supposedly shared physical structures. The concept of temperature is delineated in the first place by the Zeroth law of thermodynamics: if two things are in mutual thermal equilibrium, and one is in thermal equilibrium with a third, then the other is also in thermal equilibrium with the third. And this law allows for solids, liquids, gases, plasmas, and vacua to be in thermal equilibrium with one another. Thermal equilibrium obtains when there is no net energy transfer arising from thermal contact.¹⁷

In each instance, the property of having a temperature T , defined in terms of the foregoing equivalence relations, will be *realized* by some other property. In a gas the realizing property will be something like having n number of molecules per unit volume moving with combined kinetic energy E . In solids the realizing property will be something like having n^* number of molecules per unit volume moving with combined vibration energy E^* . And so on for plasmas, vacua etc. However, it would be wrong to *identify* temperature with these realizations. Because temperature has already been identified as that property which two

17. "Losing Your Mind: Physics, Identity and Elimination", op. cit., pp. 7-8. Note that the notion of a 'thing' here must be given a broad enough interpretation for vacua to count as things.

systems have in common when they are each capable of thermal equilibrium with a third. Specification of the property of having temperature is given by specification of the foregoing sorts of equivalence relations. And there is then no need to say anything more about what temperature is, though one can always say a great deal more about the forms of energy etc. responsible for it.¹⁸

The basic pattern which has emerged in the case of temperature is one of *unification*. The concept of temperature is a concept which covers different forms of thermal energy; it covers differences of realization. What is important here is the generality with which Blackburn claims this pattern holds.¹⁹ He claims that this pattern of unification, the emergence of concepts which cover differences of realization, is a general feature of the scientific enterprise. Consequently, the properties invoked by the explanatory theories of physical science are not constitutional properties: they are not individuated in terms of constituent physical structure. And given the foregoing example of temperature, this claim does seem plausible. For it seems to be the case that for most properties of physical science - the various forms of energy, centres of gravity, momentum, ... even such foundational properties such as uniform motion in a straight line - parallel considerations to those arising in the case of temperature show that these properties are not individuated in terms of constituent

18. "Losing Your Mind: Physics, Identity and Elimination", op. cit., p. 8.

19. "Losing Your Mind: Physics, Identity and Elimination", op. cit., p. 8.

physical structure. This is not the place to enter into extended discussion of the principles involved here. The important point is that if Blackburn's theory is correct, then it lends considerable support to the distinction between causal power and causal relevance.

Temperature, as defined by the Zeroth law of thermodynamics is an equivalence relation: it is what two systems have in common when each are capable of being in thermal equilibrium with a third. As an equivalence relation, the property of having temperature abstracts across physical structures. Because the property of having temperature is, in this sense, abstract, it is not the sort of property which has causal power. A property of standing in such and such a relation to another property does not, intuitively, seem to be the sort of thing that could bestow on things which possess it dispositions to enter into token causal relations. On the other hand, it is undeniable that a property such as having temperature enters into causal explanations; the property has causal *relevance*.

Now consider the sorts of properties which realize the property of having temperature. In a particular instance, the property, possessed by an object, of having a certain temperature T might, in that object at that time, be realized by the property, instantiated by the object, of having such and such molecules moving in such and such a way.²⁰ And it

20. Notice, one cannot say, 'moving with such and such kinetic energy'. If Blackburn's arguments are correct, kinetic energy must be regarded as an abstract, i.e. non-Tractarian, property, and, as such, not a bearer of causal power.

is this latter property which is the sort of property that could bestow upon things which instantiate it dispositions to enter into token causal relations. It is this latter property which could have causal power.

Blackburn's argument, if correct, can be used to support the distinction between causal power and causal relevance. And the generality of Blackburn's conclusion here is what is particularly striking. For it shows that a substantial proportion of physical science is concerned not with causally powerful properties, but with causally relevant properties. In the course of physical science, concepts emerge which abstract across properties which have causal power.

If the above remarks are correct, then both properties of number and what I have called scientific explanatory properties have causal relevance but do not have causal power. Therefore, even in the domain of scientific explanations, possession of causal power by a property is not a necessary condition of that property possessing causal relevance. Causal power and causal relevance are distinct properties. The primary purpose of the above arguments is to demonstrate this distinctness. However, they are also indicative of another, subsidiary, but important distinction. The distinction is between *possession* of causal power, and *taxonomization according to* causal power. A property can be taxonomized according to causal power even if it is not a possessor of such power. The distinction is highlighted by the case of scientific explanatory properties.

Consider, again, the property of having a given temperature, say T. This property, it has been argued, consists essentially in some form of equivalence relation which can hold between two or more systems. As such, the property is a possessor of causal relevance, but not of causal power. Nevertheless, it can be argued, though the property does not *possess* causal power, it is *taxonomized according to* causal power.

Consider an object x which has the property of having temperature T. Possession by x of this property is realized, in any particular case, by x also possessing some property of molecular structure, where this latter property is the possessor of causal power. What will be crucial to this underlying property if it is to realize the property of having temperature T is that it possess causal powers appropriate to such realization. Although the property of having temperature T may, in distinct cases, be realized by distinct properties, and, in this sense, is physically abstract, each realizing property of the property of having temperature T must be a property which possesses causal power appropriate to such realization.

When such a condition holds with respect to the realization of a property, I will say that the property can be *weakly taxonomized* according to causal power, or, is susceptible to *weak* causal power taxonomization. The characterization can be sharpened into the following:

A property F (where F is not a possessor of causal power) is capable of Weak Causal Power Taxonomization if and only if:

(i) If property F is possessed by object x, then, necessarily, there is some property G such that (a) x possesses G, and (b) G realizes F.

(ii) G possesses causal power P.

(iii) For all $G = \{ G_1, \dots, G_n \}$, if G realizes F, then, necessarily, G possesses P.

(iv) In each particular instance, each G is the sole realizing base of F. G is not a proper part of a realizing base of F.

The purpose of (iv) is to ensure that only causally powerful properties form realizing bases of F. It is to rule out situations in which the realizing base of F is composed of more than one property, for example, [G + H + ...] where not all of these properties have causal power. One can put the point by saying that the type of realizing base implicated in the above characterization is an *exhaustive* causally powerful property. The characterization of weak causal power taxonomization, then, amounts to necessary realization by an exhaustive causally powerful base.

The qualification 'weak' implies that there can be a stronger form of causal power taxonomization. The idea behind this stronger form is as follows. Not only is the possession of F by x realized by x's simultaneous possession of property G with appropriate causal power, but also, the possession of causal relevance R by property F is realized by possession of causal power P by property G. The causal relevance of F is realized by the causal power of G. That is:

A property F (where F is not a possessor of causal power) is capable of Strong Causal Power Taxonomization if and only if:

(i) If property F is possessed by object x, then, necessarily, there is some property G such that (a) x possesses G, and (b) G realizes F.

(ii) G possesses causal power P.

(iii) For all $G = \{ G_1 \dots G_n \}$, if G realizes F, then, necessarily, G possesses P.

(iv) In each particular instance, each G is an exhaustive causally powerful property. That is, each G is an exhaustive bearer of P.

(v) F possesses causal relevance R.

(vi) Causal power P realizes causal relevance R.

As the above characterizations makes clear, strong causal power taxonomization implies weak causal power taxonomization, but not vice versa. The strong form will occur occasionally in this thesis, but for the most part, discussion will be confined to the weak form of taxonomization. It is this form which more accurately reflects the notion of causal power taxonomization as this appears in the philosophy of mind.

The emphasis, so far, has been on stating, defending, and eliciting various corollaries of four distinctions. The remainder of this chapter will be concerned with applying these distinctions to mental properties, to mental causation and explanation, and subsequently to the principle of methodological solipsism in psychology.

There is a theory of mental properties, now widely accepted, which provides the framework for discussion in this paper. I do not endorse the theory as it stands. Rather, I believe that statements of this theory are sufficiently ambiguous to admit of a variety of interpretations, and the arguments developed in the following pages will provide the materials for the rejection of at least some of these interpretations.

The theory is that mental properties are inherently dual aspect; they essentially have two components which are distinct and independent. This theory will be termed *The Dual Component Theory of Mental Properties*. An extended quotation from McGinn makes the point:

" our intuitive conception of belief-content combines two separable components, answering to two distinct interests we have in ascriptions of belief. One component consists in a mode of representation of things in the world; the other concerns itself with properly semantic relations between such representations and the things represented. I want to suggest that the former component is constitutive of the causal-explanatory role of belief, while the latter is bound up in our taking beliefs as the bearers of truth. We view beliefs *both* as states of the head explanatory of behaviour, and as items possessed of referential truth-conditions."²¹

That is, briefly, mental properties or states necessarily have two components; what will be called the *internal* component, which

21. Colin McGinn; "The Structure of Content", in A. Woodfield, ed., Thought and Object, (Oxford, Oxford University Press, 1982), pp. 207-258, p. 210.

consists in those properties of mental states that are constitutive of their causal-explanatory role, and what I shall call the 'external' component, which consists in the referential truth-conditional properties of mental states.²²

I will understand the dual component theory in terms of the earlier introduced notion of an order of a property. If ordinary mental states are of order n , then the internal and external components of any given mental state will be understood as properties of order $(n+1)$. That is, internal and external components will be understood as properties which attach to mental properties. Talk of 'components', then, is a rather lazy way of referring to orders of properties. Just as with mental states, internal and external components can have tokens.

In order to properly understand the dual component theory, it is essential to see it in the context of the pressures on the notion of belief-content which gave rise to it. These pressures were generated by the following types of case.

22. The Dual Component Theory has done much to shape contemporary discussion in the philosophy of psychology. Some form of this theory is presupposed by Fodor's principle of Methodological Solipsism. (See Jerry A. Fodor, "Methodological Solipsism considered as a Research Strategy in Cognitive Psychology", in The Behavioural and Brain Sciences, vol. 3, no. 1. (1980). Reprinted in Fodor, Representations, Massachusetts, MIT Press, 1981) pp. 225-53. Some form of the theory is also presupposed by Stich's *autonomy principle*. See Stephen Stich; From Folk Psychology to Cognitive Science: The Case Against Belief, (Massachusetts, MIT Press, 1983). The most detailed exposition of this theory is to be found in McGinn; "The Structure of Content", op. cit. The theory is also defended in N. Block, "Advertisement for a Semantics for Psychology", Midwest Studies in Philosophy, vol. X (1986), pp. 615-78.

1. *Putnam's Twin-Earth Case.* Putnam's now famous example was first put forward in "The Meaning of 'Meaning'".²³ We are to conceive of a near duplicate of our planet Earth, called '*Twin-Earth*'. Except for certain features about to be noted, Twin-Earth duplicates Earth in every detail. The physical environments look, and largely are, identical. Moreover, many inhabitants of Earth have duplicate counterparts on Twin-Earth. These counterparts are identical with their corresponding Earthlings in point of macro/microphysical constitution. They also share with their Earth counterparts identical experiential and dispositional histories, where these are specified non-intentionally. The key difference between the two planets can be explained thus: the liquid on Twin-Earth that runs in rivers and taps is qualitatively identical with the liquid which we, on Earth, refer to with the term 'water'; that is, it is indistinguishable from the latter by any casual test. Indeed, the Twin-Earthlings refer to their liquid using the term 'water'. However, the substance that they refer to using that term is not water; it does not denote a substance whose essence lies in a structure consisting of two parts of hydrogen to one part of oxygen. Rather, it refers to a liquid which has a radically different physical structure - XYZ. Therefore, water on Earth is not the same substance as what is denoted by the term 'water' on Twin-Earth. Despite being qualitatively identical, water, and what goes by that name on Twin Earth - call it the substance retaw - are distinct substances.²⁴

23. Hilary Putnam; "The Meaning of 'Meaning'", in K. Gunderson, ed., Language, Mind and Knowledge: Minnesota Studies in the Philosophy of Science, vol. VII. (Minneapolis, University of Minnesota Press, 1975).

24. The term 'retaw' I owe to Colin McGinn (in conversation).

Suppose Herbert₁ is an English speaker of Earth, and Herbert₂ is his counterpart on Twin Earth. Neither knows the physical structure of the substance which he calls 'water'. We can also make the following assumptions about the two Herberts. Firstly, we can suppose that they are identical in point of physical constitution.²⁵ Secondly, we can suppose that they have the same inner functional states, the same behavioural dispositions, and that they exhibit the same bodily movements; where *all of these are non-intentionally specified*. When functional states, behavioural dispositions, and bodily movements are given a non-intentional interpretation, let us say that they have been *minimally specified*. Given a minimal reading, functional states, behavioural dispositions, and bodily movements supervene on physical constitution. Hence, the second supposition is a consequence of the first. Thirdly, we can suppose that the counterparts have the same minimally specified perceptual intake and streams of consciousness. In these latter cases, the notion of minimal specification is approximated by the notion of qualitative identity.

As Putnam points out, the form of words 'water is wet' means something different in the mouth of Herbert₁ than it does in the mouth of Herbert₂. The former's utterances of 'water' refer to water. The latter's utterances of the same phonetic form refer to retaw. Hence the

25. To object to this claim on the grounds that humans are largely composed of water would be to miss the point which is strictly conceptual. In any case, the example could easily be altered in terms of some substance which does not enter into the composition of human beings (aluminium, for example).

utterances differ in meaning because they differ in reference. And the form of words 'water is wet', as used on the two planets, has a different meaning in each case, because its truth-conditions vary from planet to planet. This indicates that the physical, phenomenological, and minimally specified functional and behavioural properties of a speaker do not suffice to determine speaker meaning.

Furthermore, as Putnam does not point out, the differences in meaning affect oblique occurrences in 'that'-clauses which specify the contents of mental states. Herbert₁ believes that water is wet whereas Herbert₂ believes that retaw is wet. The contents of Herbert₁'s and Herbert₂'s beliefs differ while every feature of their non-intentionally described physical, behavioural, functional and experiential histories remains the same. The difference in the contents of their beliefs, and hence in their beliefs, seems to be a product of differences in their physical environments.

2. *Burge's Counterfactual Cases.*²⁶ Putnam's Twin Earth case can be regarded as a species of *actualist* argument since it relies on comparisons between two individuals who inhabit the same possible world.

26. Tyler Burge; "Individualism and the Mental", in Midwest Studies in Philosophy vol. IV: Studies in Metaphysics, (1979), eds., P.A. French, T.E. Uehling, and H.K. Wettstein. (Minneapolis, University of Minnesota Press), pp. 73-121. Also, "Two Thought Experiments Reviewed", in Notre Dame Journal of Formal Logic, vol. xxii, pp. 284-93. "Individualism and Psychology", in The Philosophical Review, vol. XCV, (1986), pp. 3-45. "Cartesian Error and the Objectivity of Perception", in P. Pettit and J. McDowell eds., Subject, Thought and Context (Oxford, Oxford University Press), pp. 117-36. The counterfactualist status of Burge's arguments are not essential to them. The arguments could easily be re-run in the actualist framework familiar from the work of Putnam.

Tyler Burge has put forward various arguments which rely on comparison of a single individual in various possible worlds, that is, in various counterfactual situations. Burge's thought experiments are designed to show that physical/minimal functional/experiential duplicates would differ in beliefs as a result of differences in the way the subject's linguistic community applies terms in the language.

In his leading case we are to imagine a person, call him Oscar, who has a number of psychological states which are commonly attributed with content clauses containing the term 'arthritis' in oblique occurrence. For example, he believes that he has had arthritis for years, that the arthritis in his fingers is more painful than the arthritis in his toes, he desires that the arthritis should cease, and so on. However, in addition to these beliefs, which may be correct, he also believes he has developed arthritis in his thigh. However, arthritis is an ailment which affects only the joints. Therefore, Oscar's belief that he has arthritis of the thigh is a false belief. What is important here is that, despite his misapprehension, we still regard Oscar's belief as a belief about arthritis. It is simply a false belief about arthritis. The principle underlying our intuition in this case seems to be the following: if other people in his language community can bear attitude A (belief, desire etc) towards the proposition that arthritis is F, so too can Oscar. Of course, Oscar differs from other members of his language community in that much of what he believes about arthritis is false. However, his false beliefs about arthritis are, nevertheless, attitudes directed toward arthritis.

We now imagine Oscar in a counterfactual situation.²⁷ In this situation Oscar possesses the same physical and, hence, minimal functional and behavioural states or dispositions which he actually has. Furthermore, Oscar's non-intentional phenomenological experience is the same - he has the same pains, visual fields, images, streams of consciousness, etc. The counterfactuality in the situation touches only on Oscar's linguistic environment. In the counterfactual case, the phonetic form 'arthritis' - as used by physicians, lexicographers, and competent laymen - applies not only to arthritis (in the sense of an affliction of the joints), but also to various other rheumatoid ailments. That is, the standard counterfactual use of the term is to be conceived to encompass Oscar's actual misuse. In this case, our intuition tells us that Oscar, in the counterfactual situation, does not have arthritis attitudes. That is, counterfactual Oscar lacks the attitudes which we, in the actual world, attribute with content clauses containing the term 'arthritis' in oblique occurrence. It would be wrong for us, as speakers of English, to say of counterfactual Oscar that his beliefs, desires etc., are ever directed to a proposition of the form '... arthritis ...'. For Oscar in the counterfactual situation, unlike Oscar in the actual situation, does not have the concept of arthritis; he has the concept of counterfactual arthritis (= arthritis + rheumatism, as we might say).

27. Use of the expression 'Counterfactual Oscar' is simply meant to record the fact that Oscar is in a counterfactual situation. It is not meant to imply, for example, that Oscar is a different person in each situation. Questions concerning the transworld stability of individuals are beyond the scope of this thesis. Such stability is assumed throughout. The arguments could easily be revised if this stability ultimately turns out not to obtain.

The argument here does not turn on any supposed natural kind status of arthritis. Arthritis is, in fact, not a natural kind. It covers distinct ailments - rheumatoid arthritis, osteo-arthritis, gout etc. - which have widely different causes and etiologies. Moreover, Burge-type arguments can be constructed for kinds which are patently non-natural - brisket, sofa, chicory etc. Indeed, Burge claims that the relevant types of argument can be constructed for "... any notion that applies to public types of objects, properties, or events that are typically known by empirical means."²⁸

The differences in the content of the attitudes involved stem from differences in linguistic usage. The differences in content are again the result of differences in the reference of the terms used in the expression of that content, but in this case the differences in reference derive from differences in linguistic usage, not from differences in the physical environment. Because the differences in content result from divergences in the reference of terms, these content differences are still due to the external component of mental states as I have characterized that notion.

3. *Indexicals*. The third category of cases motivating the distinction between internal and external components of mental states derives from certain treatments of indexical elements in thought and language. These interpretations are not uncontroversial, and this is not the place to

28. "Individualism and Psychology", op. cit., p. 6.

become embroiled in controversies surrounding their relative merit.²⁹ I will list them for the sake of completeness, showing how they support the internal/external distinction, and without addressing the question of their relative merit.

In addition to classical indexical beliefs, the category of relevant mental states here includes demonstrative thoughts and de re thoughts about individuals. On certain treatments of indexicality it is recognized that indexical beliefs can be ascribed by content sentences which have, in Kaplan's terminology, the same *character*, but which differ in referential truth-conditions. [Kaplan's notion of character is roughly equivalent to Frege's notion of sense. However, there is even less inclination on my part to become involved in the controversies surrounding the interpretation of this latter notion. Hence, the less controversial 'character'].³⁰ For example, the sentence 'today is fine', while expressing the same character on each occasion of use, differs in point of referential truth-conditions on each day in which it is used. Similarly, the utterance 'I am tired' possesses the same character on each occasion of use, but differs in referential truth-conditions depending on the utterer and the time of utterance. Thus

29. Principle proponent of the outlined view is J. Perry: "Frege on Demonstratives", in The Philosophical Review, vol.86 (LXXXVI) (1977), pp. 474-97. "The Problem of the Essential Indexical", in Nous vol.13, (1979), pp. 3-21. Also influential here has been David Kaplan's unpublished monograph "Demonstratives". Principle opponent is Gareth Evans: "Understanding Demonstratives", in H. Parrett and J. Bouveresse eds., Meaning and Understanding, (Berlin, New York, Walter De Gruyter, 1980).

30. The expression 'character' is due to Kaplan, "Demonstratives", op. cit.

suppose that I and my Twin Earth Doppelganger both have a belief specified by the content sentence 'I am tired'. Again, we can assume identity of physical, minimal functional, dispositional, and experiential types. Then the beliefs had by us both have the same character, but differ in referential truth-conditions: the truth-conditions of my belief are that I am tired; the truth-conditions of my Doppelganger's belief are that he is tired. Similarly with *de re* thoughts about individuals. For example, suppose Twin Earth is just like Earth except that my wallet is in my coat pocket while my doppelganger's wallet is not in his coat pocket. I believe (truly) that my wallet is in my coat pocket. My Doppelganger has the counterpart belief. His belief is false, mine is true. And, crucially, the beliefs differ in referential truth-conditions. The truth conditions of my belief are that my wallet is in my pocket, the truth-conditions of his belief are that his wallet is in his pocket.³¹ The Doppelganger is identical with me in point of physical, minimal functional, dispositional, and experiential types. Nevertheless, the contents of our beliefs, and hence our beliefs, differ. Beliefs are not in the head.

These three types of case, if cogent, indicate a need to distinguish two distinct components of mental properties. One component, which I have called the *internal* component, is that component concerned with the production of action, and has causal powers relevant

31. This example is borrowed from Daniel Dennett; "Beyond Belief", in A. Woodfield, ed., Thought and Object (Oxford, Oxford University Press, 1982), p. 12.

to such production. Thus, the internal component of a mental state is, basically, the causal or functional role which that mental state plays in an agent's psychology. The internal component is identical with what many have called the mental state *narrowly individuated*, or the *narrow mental state*.³² The other component, the *external* one, consists in the referential truth-conditional properties of mental states: that is, the properties possessed by any given mental state in virtue of which it represents a portion of the world. The internal and external components taken together make up the mental state *widely individuated* or the *wide mental state*.

The issue is perhaps clearest with indexicals. The bifurcation of action-guiding and truth-conditional components of mental states is advocated by Perry.³³ What I have called the internal component or narrow mental state, Perry, deliberately echoing Frege, calls the sense of that state. What I have called the wide mental state, Perry calls the thought. He writes:

" We use senses to individuate psychological states in explaining and predicting action. It is the sense entertained, and not the thought apprehended, that is tied to human action. When you and I entertain the sense of 'A bear is about to attack me', we behave similarly. We both roll up in a ball and try to be as still as possible. Different thoughts apprehended, same sense entertained, same behaviour. When you and I both apprehend the thought that I am about to be attacked by a bear, we behave differently. I roll up in

32. Putnam, "The Meaning of Meaning", op. cit. Also Fodor, "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology", op. cit. Burge, "Individualism and Psychology", op. cit.

33. "Frege on Demonstratives", op. cit.

a ball, you run to get help. Same thought apprehended, different sense entertained, different behaviour."³⁴

For another example, suppose I possess the thought that today is beautiful. Then, due to the indexical functor 'today', the truth-conditions of my thought depend crucially (but imperceptibly to me) on such events as the moving of a clock at the Greenwich Observatory. But it is wildly implausible to suppose that the portion of my mental state which is responsible for my behaviour depends not just on my internal constitution at the time, but also on such causally remote features as the disposition of the parts of some official time-keeper.³⁵

Consider again Putnam's Twin Earth case. In terms of the role played in the agent's psychology, that is, in terms of the dispositions to behaviour it will bring about, Herbert₁'s belief, attributable by the phonetic form 'water is wet', is type-identical with the belief of Herbert₂ attributable by the same phonetic form.³⁶ However, the beliefs differ with regard to their referential truth-conditions. The truth-

34. "Frege on Demonstratives", op. cit., p. 494.

35. The example is borrowed from Dennett, "Beyond Belief", op. cit. The example should perhaps be recast in terms of the rotation of the Earth with respect to the sun in order to account for the possibility of having the thought before the building of the observatory.

36. It is sometimes denied that the two states are identical in terms of the dispositions to behaviour they produce. It is argued, for example, that the former state will bring about the drinking of water, while the latter state will bring about the drinking of retaw. I will adopt the standard view that the two states are identical in point of the dispositions to behaviour they produce. This claim is defended by Fodor in Psychosemantics (Massachusetts, MIT Press, 1987), ch.2, especially pp. 34-41. Also, Frank Jackson and Philip Pettit, "Functionalism and Broad Content" (Unpublished).

conditions of the former are that water is wet, those of the latter are that retaw is wet. That is, the beliefs differ with regard to, and in virtue of, their external component. Thus, although their functional roles are identical, the beliefs are regarded as distinct. Similarly with the beliefs of actual and counterfactual Oscar. From the point of view of functional role, the two beliefs attributed by the single form 'I have arthritis in my thigh', are type-identical. However, they differ with regard to referential truth-conditions, and are hence regarded as type-distinct.

The above types of case, then, are indicative of a need to distinguish internal and external components of mental states. This is because they show that the external component of a belief is not always or necessarily encoded in the internal component: the causal or functional role of a mental state does not determine its truth-conditions. In all three cases, the internal components are fixed while the external components vary. Therefore, internal components do not determine external components.

Moreover, it is easy to think of examples where failure of determination obtains in the converse direction. Such examples include those cases of non-trivial identity statements which exercised Frege.³⁷ The Morning Star is identical with the Evening Star, but Jones, who believes that the Evening Star can be seen low in the western sky

37. "On Sense and Reference". Reprinted in Translations From the Philosophical Writings of Gottlob Frege, ed., P. Geach and M. Black, (Oxford, Blackwell, 1957), pp. 56-78.

shortly after sunset, may not believe the same of the Morning Star. Here we have a case of two beliefs with the same truth-conditions but with a different cognitive significance, and, hence, a different role in the agent's psychology. That is, we have a case of two beliefs which are type-identical with regard to their external component, but which are regarded as distinct beliefs because of the type-distinctness of their internal components.

These types of cases, the former which show that internal components do not determine external components, and the latter which show that external components do not determine internal components, thereby make a case out for distinguishing the two types of component. Given this distinction, I will now go on to discuss how the distinction between causal power and causal relevance applies to it. The results obtained here will be used in an examination of the role played by the principle of methodological solipsism.

3

The claim that will be argued for in this section is that standard formulations of the dual component theory house an important ambiguity, an ambiguity which can be brought out only when the distinction between causal power and causal/explanatory relevance is properly understood. Consider, for example, the following passage from McGinn:

" consider factive propositional attitudes, e.g. knowing, remembering, perceiving. We do commonly employ these in explanatory

contexts, yet it would be agreed that they are hybrid states requiring the world as well as the agent's head to be a certain way. What we should say of this is clear: *only the internal component of the condition reported is doing explanatory work - the rest is, from an explanatory point of view, idle.*"³⁸

What is crucial here is that McGinn locates *explanatory relevance* in the internal component of mental states. And given that McGinn regards psychological explanation as a species of causal explanation, the above passage can be interpreted as expressing the claim that *internal components are the loci of causal relevance*. That is, it is internal components which play an explanatory role in causal explanations. And this, in turn, entails that statements of these explanations will contain terms which *name* or *make* some similar form of *direct reference* to the internal components in question. However, in the very next sentence, McGinn goes on to make a claim which can (but need not) be given a somewhat different interpretation:

" Since I similarly hold that beliefs have hybrid content I can take a similar line: *the explanatory force of the content ascription attaches only to the contribution the words in the content clauses make in their capacity as specifiers of internal representations.*"³⁹

The claim expressed in the first passage entails the claim expressed in the second, but not vice versa; the claims are distinct. For it may be true that content ascriptions have causal/explanatory force only because they, perhaps indirectly, specify internal components of mental states,

38. "The Structure of Content", op. cit., p. 215. Emphases are mine.

39. "The Structure of Content", op. cit., p. 215. Emphases are mine.

but false that statements of the explanations relate or contain terms which name internal components.

The distinction between causal power and causal relevance indicates why this is so. The first passage clearly identifies internal components as possessors of explanatory relevance, and, given that the explanations in which they figure are causal, this means that internal components are the loci of causal relevance. Now, if internal components possess causal relevance, then, by the definition of that notion, they must play an explanatory role in causal explanations. But, if they are to play such a role, then statements of those explanations must contain terms which name or in some other way directly refer to those internal components. However, the second passage claims only that the explanatory force of content-ascriptions is ultimately made possible by the internal components. And this in no way entails that internal components figure directly in the relevant explanations, or that statements of those explanations contain terms which name internal components. For it might be that the explanatory force of content-ascriptions is made possible by the fact that internal components have causal *power*. And this does not, given the distinctness of the two properties, entail that internal components have causal relevance also.

The interpretation of the dual component theory as claiming that internal components have causal (and hence explanatory) relevance, I will call the *explanatorily narrow* interpretation of that theory. An interpretation which identifies internal components as loci of causal power, but not of causal relevance, I will call an *ontologically narrow*

interpretation. Any plausible interpretation of the dual component theory must be ontologically narrow, but, it will be argued, need not be explanatorily narrow.⁴⁰

It will be seen in the following section that these two possible interpretations of the dual component theory give rise to two possible interpretations of the principle of methodological solipsism.

4

The apparatus provided by the dual component theory can be used to characterize a notion of *externalism* with regard to mental states and mental content which will be used throughout this thesis. The three types of case motivating the dual component theory are indicative of the

40. The distinction between explanatory and ontological narrowness can be used to answer an important objection to the dual component raised by Hornsby in "Physicalist Thinking and Conceptions of Behaviour", in P. Pettit and J. McDowell eds., Subject, Thought and Context op. cit. On page 111 she writes:

" ... it is hard to see how anyone is in a position to claim that there are states whose ascription is explanatory of their behaviour unless he can demonstrate that the ascription of such states does, or would, cast light on behaviour. We know of course that such states as we ascribe - beliefs having content, for instance - do cast light. But it is no help then to be told that there must be states which lurk behind the states we ascribe and which carry their explanatory force. It is a strange idea that the satisfaction yielded by common sense explanations has its source in something to which the parties are quite ignorant ... "

Hornsby's objection is important and correct, but only applies to an explanatorily narrow interpretation of the dual component theory. It is the burden of this thesis to show that the explanatorily narrow interpretation must be rejected in all its forms.

external individuation of (certain sorts of) mental properties. That is, the cases indicate that what mental states a person possesses is dependent on the nature of that person's - physical and social - environment. More precisely, the principle of externalism may be characterized as follows:⁴¹

Externalism is the thesis that an individual's mental state types cannot, even in principle, be individuated independently of the natures of the empirical objects, properties and relations which are external to that individual. There is a necessary individuating relationship between the mental state types possessed by an individual and the external objects, properties and relations to which those states are connected by the relation of reference.

As this definition makes clear, the thesis of externalism only applies to states which have referential or representational properties; it only applies, that is, to *intentional* mental states.

The principle of externalism can be interpreted in terms of the dual component theory:

Externalism is the thesis that there is a necessary individuating relationship between the mental state types possessed by an individual, and the external objects, properties and relations to

41. c.f. Burge, "Cartesian Error and the Objectivity of Perception", *op. cit.*, pp. 118-19.

which those states are related in virtue of their external components.

It is a corollary of this claim that assumes principle importance in this thesis. I will call this corollary the principle of externalism*:

Externalism*: The external component of any intentional mental property is an essential type-individuating feature of that property.

Within this characterization, a distinction can and should be drawn between what I shall call *total* and *partial* externalism*.

Total Externalism*: The external component of any intentional mental property is the *only* essential type-individuating feature of that property.

Partial Externalism*: The external component of any intentional mental property is an essential type-individuating feature of that property, but is not the only essential type-individuating feature of the property.

The distinction between total and partial externalism* will assume importance later.

Solipsism, or individualism, with respect to mental states can be understood as the converse of externalism. Solipsism is the view that

there is no necessary individuating relationship holding between an individual's (intentional) mental properties and the external objects, properties, and relations to which those properties are referentially related.⁴²

The principle of *Methodological Solipsism* will be understood here as a principle which purports to constrain theory formation in cognitive science. As such, it makes no claim about how mental states are individuated in folk psychology, but only about how they *should* be individuated for the purposes of cognitive science. In recent discussions, the principle of methodological solipsism appears in several guises: As the claim that psychology should be exclusively concerned with psychological states in the narrow sense.⁴³ As the claim that psychology is committed to viewing psychological processes as computational (hence as formal operations defined over representations).⁴⁴ As the claim that the states and processes that ought to be of concern to the psychologist are those that supervene on the current, internal, physical state of the organism.⁴⁵ As the claim that psychology is committed to individuating mental states in terms of their causal powers.⁴⁶

42. c.f. Burge, "Cartesian Error and the Objectivity of Perception", op. cit., pp. 118-19.

43. Putnam, "The Meaning of 'Meaning'", op. cit.

44. Fodor, "Methodological Solipsism considered as Research Strategy in Cognitive Psychology", op. cit.

45. Stich, From Folk Psychology to Cognitive Science: The Case Against Belief, op. cit., ch.8.

46. Fodor, "Individualism and Supervenience", op. cit. Equivalently, Psychosemantics, op. cit., ch.2.

At the core of all these formulations is what one might call *the Causal Objection* to folk psychological locutions. More precisely, the various arguments appearing in current literature which motivate the demand for methodological solipsism are variants of a single Ur-argument which runs something like this:

- (1) The goal of respectable scientific psychology is the causal explanation of behaviour. To this end it aims to develop causal generalizations and explanations.
- (2) Causal generalizations and explanations subsume events in virtue of their causal properties.
- (3) Therefore, the taxonomy of psychological states demanded by a respectable scientific psychology is a taxonomy based around causal powers.
- (4) Certain properties of mental states - the referential truth-conditional properties - are not determinants of their causal powers.
- (5) Therefore, referential truth-conditional properties must play no role in respectable psychological taxonomization.⁴⁷

And, it turns out that what is left of the mental state when you subtract the referential truth-conditional properties is the *narrow* mental state, the *computational* state, that portion of the mental state which *supervenies on current, internal, physical, states of organisms*, or

47. The argument is perhaps most explicit in Fodor, "Individualism and Supervenience", *op. cit.* Equivalently, Psychosemantics, *op. cit.*, ch.2.

that portion of a mental state which is determinant of its causal powers.

It is easy, then, to reinterpret this methodological constraint in terms of the internal/external distinction previously developed. Reinterpreted, the constraint can be expressed thus: cognitive psychology is to be concerned purely with the internal component of mental states. The external component lies beyond its scope. The external component - the referential truth-conditional component - has no causal powers with respect to the production of action.

However, the argument expressed in (1)-(5) leads to a conclusion which is ambiguous. And the ambiguity stems from the fact that there are two possible interpretations of the dual component theory. In the above formulation, the ambiguity is reflected in the unclarity of the notion of psychology being 'concerned with' the internal component. The notion can be explicated in two distinct ways. On the one hand we might interpret it as the claim that cognitive science seeks to develop laws, generalizations, and explanations which relate internal components of mental states. In this case, statements of those laws, generalizations, and explanations must contain terms which *name* or *make* some form of *direct reference* to those internal components. In other words, this interpretation would *make* internal components possessors of causal (and hence explanatory) *relevance* as well as of causal *power*. On the other hand, one might interpret the constraint of methodological solipsism as the claim that cognitive science seeks laws, generalizations, and explanations which hold between mental states *in virtue of* their

internal components. But this claim need attribute only causal power to internal components; it need not attribute causal relevance to them as well. For, given the distinction, we can allow that the laws, generalizations, and explanations of cognitive psychology are made possible by internal components - and which, therefore, hold *in virtue of* internal components - without also claiming that these internal components figure *directly* in those laws, generalizations and explanations. That is, we can allow that the laws, generalizations and explanations of cognitive psychology hold *in virtue of* internal components, because internal components have causal *power*, without claiming that they hold *between* internal components, without claiming, that is, that the internal components have causal relevance as well as causal power.

The ambiguity expressed above indicates a need to distinguish two forms of methodological solipsism:

- (1) *Ontological Methodological Solipsism* (OMS). The laws, generalizations, and explanations of cognitive psychology are ultimately made possible by the internal components of mental states, because it is these components which have causal power. However, according to OMS, internal components do not also have causal relevance in addition to causal power.
- (2) *Explanatory Methodological Solipsism* (EMS). EMS agrees with OMS that the laws, generalizations and explanations of cognitive psychology are made possible because internal components have causal power. However, it also makes the far stronger claim that the

internal components (or some properties thereof) figure directly in these laws, generalizations and explanations. That is, internal components (or some properties thereof) possess causal relevance as well as causal power.

As should be clear, these two forms of methodological solipsism are a direct result of applying the distinction between causal power and causal relevance. One of the principle reasons, I believe, for the consistent failure to recognize this distinction, is insufficient attention being paid to the use of the phrase 'in virtue of'. As the above considerations should make clear, this phrase is importantly ambiguous. Hence, it will not be used in future discussion.

It is important to remember that causal/explanatory relevance is always relative to a domain of explanation. Hence, the principle of explanatory methodological solipsism is properly understood as making a claim about causal/explanatory relevance in cognitive psychology. That is, it is to be understood as claiming that it is internal components, or properties thereof, which possess causal/explanatory relevance *of a sort appropriate to scientific cognitive psychology*. Note that this is compatible with wide mental states possessing causal/explanatory relevance of a sort appropriate to folk psychology. This qualification remains in force throughout the thesis; the type of causal/explanatory relevance at issue will be the type appropriate to cognitive, rather than folk, psychology.

Within explanatory methodological solipsism, a further distinction can and should be drawn.⁴⁸

(a) *Formal Explanatory Methodological Solipsism* (FEMS). This claims that the causally relevant properties possessed by internal components are formal or syntactic. Hence, the explanations of cognitive psychology will relate these formal/syntactic properties, and statements of these explanations will contain names of those properties.

(b) *Conceptual Explanatory Methodological Solipsism* (CEMS). This claims that the causally relevant properties possessed by internal components are not syntactic, but, rather, *conceptual*, or, properties of (some form of) content. These conceptual properties are constitutive of what has become known as *narrow content*. For the moment, the precise nature of these properties will be left open.

Acceptance of FEMS leads to what has become known as the *syntactic theory of the mind*. This will be the subject of chapters III-V. The implications of CEMS will be the subject of the next chapter.

48. A similar distinction is to be found in K. Bach, "De Re Belief and Methodological Solipsism", in A. Woodfield, ed., *Thought and Object*, op. cit., pp. 121-51. However, Bach does not distinguish ontological from explanatory methodological solipsism. Consequently, his distinction is not restricted to explanatory methodological solipsism. It is possible to distinguish formal and conceptual versions of *ontological methodological solipsism*, but such a distinction does no useful theoretical work.

In order to properly understand the principle of explanatory methodological solipsism, it is essential to see it in the context of the distinction between a property's possessing causal power and a property's being susceptible to causal power taxonomization. In accordance with the earlier discussion, it is the weak form of taxonomization which is at issue here.

The primary purpose of EMS is to ensure, not that the explanatory properties invoked by a scientific psychology will be possessors of causal power (though EMS would certainly not rule this out), but, rather, that they will be (weakly) taxonomized according to causal power. It is no part of EMS that the explanations of a scientific psychology should invoke only those states which have causal power. Such a claim would be as implausible here as it is in the case of physical theory.

EMS is committed only to the claim that the explanations of scientific psychology should directly refer only to those states which are (weakly) taxonomized according to causal power. Therefore, EMS amounts to the claim that any property which possesses causal relevance of a sort appropriate to a scientific cognitive psychology must either possess causal power, or be (weakly) taxonomized according to causal power.

In this thesis, the following position vis-a-vis the principle of methodological solipsism will be defended: It is possible to accept OMS while rejecting EMS. That is, it is possible to allow that the explanations of a scientific cognitive psychology are ultimately made possible by internal components - because these are the loci of causal power - but that statements of those explanations should refer to mental states widely individuated. It is such states which possess causal relevance of a sort appropriate to a scientific psychology even though they are not taxonomized according to causal power.

This type of approach faces a serious objection, the reply to which can be regarded as constituting the bulk of this thesis. The objection is a form of Occam's Razor.

The suggestion given above has the following parallel in physical theory. Suppose someone recommended that the explanations of physical theory should not invoke the property of temperature, but, rather, the property of temperature*. This corresponds to the old property of temperature plus the property of standing in some external relation, *, to some other property. This external relation in no way affects the causal relations entered into by instances of temperature. Thus, temperature is, but temperature* is not, taxonomized according to causal power. The immediate and obvious objection is this: What possible rationale could there be for not eliminating the external relation * from the explanations of a scientific physical theory. It could be granted that some cultures might employ temperature* as an explanatory property of their physical theory, but this must be regarded as a

symptom of ignorance, to be eliminated from any sufficiently advanced physical theory.

Similarly, so the objection goes, in our ordinary (i.e. primitive) folk psychological explanations we might invoke properties which are not taxonomized according to causal power. Such properties might partly be constituted by external relations such as referential truth-conditional relations. But this must be regarded as merely a reflection of our folk psychological ignorance. And once our psychology has developed to a respectable scientific level, there can be no excuse for continuing to implicate such properties in its explanations.

The remainder of this thesis is, in effect, a reply to this question. The basis of this reply is that those purported causally relevant properties of internal components cannot be identified independently of the wider semantic properties of the mental states which possess those internal components. Therefore, the sort of abstraction visualised by EMS, of properties which possess causal relevance of a sort appropriate to scientific psychology from those that do not, is not a real possibility. With the distinction between formal and conceptual explanatory methodological solipsism, this reply comes in two parts.

In chapter II, it will be argued that the only grip we have on the conceptual properties referred to in CEMS, that is, the only grip we have on the notion of narrow content, is, in any given case, provided by ordinary wide semantic content. Narrow content cannot be identified

independently of ordinary semantic content. In the remaining chapters, it will be argued that syntactic properties, as they are required by the syntactic theory of mind, cannot be identified independently of the semantic properties which they realize. Therefore, both CEMS and FEMS, and, hence, EMS *tout court*, will be rejected. This thesis is primarily about FEMS. Consequently, discussion of CEMS will be brief. Considerations of completeness aside, the rationale for the discussion of CEMS will be that the rejection of this principle exhibits a pattern which, it will be argued, is repeated in the rejection of FEMS.

Appendix I:

Dennett's Illata/Abstracta Distinction as a Case of the Causal
Power/Causal Relevance Distinction.

1

In his earlier writings, Dennett puts forward an account of mental states which is explicitly *instrumentalist*.¹ The reason for this is of some importance here, and stems from Dennett's view of the role played by rationality in psychological ascriptions.²

Dennett's, now famous, example concerns a chess-playing machine. In predicting the moves the machine will make, we can, says Dennett, adopt three possible postures. Firstly, we might predict the moves from a knowledge of the machine's physical constitution. Secondly, we might predict the moves by reference to the functional or teleological structure of the machine. Dennett says that in making predictions the first way, we are adopting a "physical stance" towards the machine, and that in making predictions the second way we are adopting a "design stance." Thirdly, we might adopt what Dennett calls an "intentional stance".

-
1. Daniel Dennett, Brainstorms (Massachusetts, MIT Press, 1978). Especially "Intentional Systems", pp. 3-22.
 2. See "Intentional Systems", *op. cit.*

Design stance predictions are riskier than physical stance predictions. The latter, predictions inferred from a conjunction of the relevant physical laws and a specification of the state of the machine at time *t*, can go wrong only insofar as the physics of the machine is not deterministic. However, design stance predictions will be reliable only insofar as the machine functions as it is supposed to function, for example, if it does not break down.

The intentional stance, like the design stance, is prone to kinds of fallibility. Adopting the intentional stance involves predicting what the system will do by working out what it would be *rational* to do, and then predicting that the system will do that. Intentional stance predictions inherit the fallibility of design stance predictions. Even if the system *is* rational, it will still be true that intentional stance predictions will go astray insofar as the system malfunctions, or insofar as the physics of the machine is not deterministic. But, in addition to this, intentional stance predictions are fallible in themselves: Even if the system does not malfunction, it may fail to do the rational thing because it was not designed to be entirely rational.

According to Dennett, the decision to adopt the intentional stance is purely pragmatic. On the one hand we may not know the physical or functional constitution of the system, but may still know what it would be rational for the system to do. On the other hand, even if we do know the physical or functional structure of the machine, predictions based on this knowledge may become unfeasible in practice when the system is sufficiently complicated. In either case, it is essentially pragmatic

considerations that force us to intentional ascriptions: in particular, it is not considerations of *truth*. Dennett says: "The decision to adopt the strategy is pragmatic and is not intrinsically right or wrong. One can always refuse to adopt the intentional stance ... one can switch stances at will ...etc."³

It is these considerations which lead Dennett to his instrumentalism: Since intentional stance predictions will work only insofar as we are dealing with rational systems, an assumption of rationality is implicit in every such prediction. And the notion of rationality, here, is an ideal or *normative* notion. But the assumption is ours to make or withhold. That is, the assumption is a heuristic device, an idealization, justifiable only insofar as it leads, by and large, to intentional stance predictions. So, Dennett infers the instrumental character of intentional states from the heuristic nature of the rationality assumption which grounds intentional ascriptions. In his early work, then, Dennett regards the states posited by folk psychology as instrumental myths, useful for the prediction of behaviour.

2

Dennett, in his later writings, abandons this instrumentalist account of folk psychology. The principle reason for this seems to be

3. "Intentional Systems", op. cit., p.7.

his interest in providing a theoretical foundation for the notions of personhood, moral agency, responsibility, and so on.⁴ According to Dennett, these notions require two things. Firstly, they require intentional concepts.⁵ Secondly, and more importantly for present purposes, they require that intentional mental states be construed as possessors of causal relevance, and, hence, are not to be construed instrumentally.

The notion of causal relevance, of course, is not Dennett's. There is, in Dennett's work, no principled distinction between causal power and causal relevance. However, I want to suggest that the distinction between causal power and causal relevance can be usefully employed to throw light on a distinction which is foundational to Dennett's later view of intentional states. The distinction is that between Illata and Abstracta.

Dennett's later account of intentional mental states is based around the illata/abstracta distinction. Illata are the ineliminable posits of a scientific theory. That theory must quantify over and, hence, be ontologically committed to anything which is to count as an illatum of that theory. Abstracta, on the other hand, are merely

4. See, for example, "Intentional Systems in Cognitive Ethology: The Panglossian Paradigm Defended", in The Behavioural and Brain Sciences vol.6, pp. 343-90. Dennett discusses moral agency in "Reply to Professor Stich", Philosophical Books, vol. XXI, pp. 73-76. Rational agency is discussed in "Mechanism and Responsibility" in Brainstorms, op. cit., pp. 233-55.

5. "Intentional Systems in Cognitive Ethology: The Panglossian Paradigm Defended", op. cit.

calculation bound entities. The example of an abstractum given by Dennett is a centre of gravity.⁶

Dennett leaves the illata/abstracta distinction extremely unclear. I want to suggest that it can usefully be understood as the distinction between those states which have causal relevance (abstracta) and those states which have both causal power and causal relevance (illata). Perhaps the best way to approach Dennett's understanding of the distinction is by looking at the example of an abstractum given by Dennett - a centre of gravity.

The first important point to notice is that talk of centres of gravity can, in any particular instance, always be in principle eliminated in favour of talk of masses and positions. The reason is that, for any object, there is a formula which will *map* the centre of gravity of that object onto a function of (i) the masses of the component particles of that object, and (ii) the positions of those particles. The formula can be represented thus:

$$X = \frac{MaXa + MbXb + McXc \dots}{Ma + Mb + Mc \dots}$$

where X is the centre of gravity, Ma is the mass of a constituent particle, and Xa is the position of that constituent particle. This

6. Dennett, "Three Kinds of Intentional Psychology" in R. Healey, ed., Reduction, Time and Reality, (Cambridge, Cambridge University Press, 1981), pp. 37-61.

formula, by mapping centres of gravity onto masses and positions provides the materials necessary for elimination of talk of centres of gravity in physical science. Hence, centres of gravity are not ultimate posits of physical theory.

The second point to notice is that it is plausible to speak of centres of gravity as possessors of causal relevance, but not as possessors of causal power. Consider, first, causal relevance. It is plausible to regard the notion of a centre of gravity as playing an explanatory in a causal explanation. For example, consider a situation in which some object - a ming vase - is propelled by some external force across a table until it falls off. Then we would regard the following as (at least part of) a genuine causal explanation.

" The vase fell off the table because it had reached a point where its centre of gravity was no longer vertically above any part of the table".

So, it seems plausible to regard centres of gravity as possessors of causal relevance. On the other hand, it does not seem plausible to regard them as possessors of causal power. Centres of gravity are, like geometrician's lines, not real entities at all. They are theoretical abstractions. A centre of gravity may, in any particular instance, be located at the same place as some real point of an object, but, for well known reasons, it cannot be identified with that point. As a theoretician's abstraction, a centre of gravity does not seem to be the sort of thing which can bestow on objects which possess it dispositions

to enter into causal relations. Similar remarks apply to other theoretical abstractions - a point charge would be an example here.

Therefore, from the discussion of centres of gravity, there appear to be two essential features possessed by abstracta:

(i) Abstracta are, in any particular case, always in principle eliminable.

(ii) Abstracta are the possessors of causal relevance, but not of causal power.

Moreover, the reason why centres of gravity possess each of these features seems to be the same: There are physical facts, viz., a certain distribution of mass, in virtue of which the centre of gravity of an object is where it is. A centre of gravity is mapped onto these physical facts by way of the formula given above, and this accounts for their eliminability. And, furthermore, it is properties such as having a certain distribution of mass which are possessors of causal power. And, so, the reason that centres of gravity possess causal relevance in the absence of causal power is because they can be mapped onto those states which do possess causal power. And we can plausibly accept this conclusion for abstracta generally. An identical pattern emerged in the earlier discussion of numbers and scientific explanatory properties.

Dennett's view of intentional states as relevantly similar to such things as centres of gravity is made explicit in the following passage,

and this passage also bears out the view of abstracta as eliminable causally relevant entities:

" My view is that no simple, direct 'reductionist' accomodation can be made ... and that the best sense can be made of folk psychology (of belief and desire talk in particular) if it is viewed *instrumentally*. So I am an 'instrumentalist' - but not a *fictionalist* ... Attributions of belief and desire are not just 'convenient fictions': there are plenty of honest-to-goodness instrumentalist truths ... Consider the truths one can assert regarding an instrumentalist entity such as a *centre of gravity*.

As you slide the lamp over the edge of the table, it will remain upright so long as the centre of gravity is located over a point on its base still on the table.

You can move the centre of gravity of the lamp down by filling its base with water, and to the side by sticking a wad of chewing gum on the side.

Are centres of gravity fictions? In one sense, perhaps, but there are plenty of true, valuable, empirically testable things one can say with the help of the term - and one doesn't fret about not being able to 'reduce' an object's centre of gravity to some particle or other of the lamp. Explanations may refer to centres of gravity. Why didn't the doll tip over? Because its centre of gravity was so low. (This explanation is not obviously 'causal' but it surely competes with others that are: 'Because it is glued to the table', 'Because it is suspended by invisible threads') ... I want to claim much the same sort of thing about belief claims."⁷

It is important to note here that Dennett's use of the term 'instrumentalist' is non-standard. The standard use of that term is occupied by Dennett's term 'fictionalist'. I will retain use of the former.

Dennett's account of *illata* is equally as obscure as his treatment of abstracta. However, the general idea seems to be something like this:⁸ We can analyse a given organism's intentional states according

7. "Intentional Systems in Cognitive Ethology ...", op. cit., p.380.

8. See "Three Kinds of Intentional Psychology", op. cit., pp. 55-57.

to the top-down strategy. The important point here is that each of the subsystems revealed by analysis will be an intentional system, albeit more 'stupid' than the systems higher up in the hierarchy. This process should culminate in low-level intentional systems, sufficiently low-level to be characterized in physical terms. It is these low-level systems which, in psychology, constitute illata. It is these illata which are the ineliminable posits of what Dennett calls 'subpersonal cognitive psychology'.⁹

The illata of subpersonal cognitive psychology just are physical states of organisms, intentionally characterized. And it seems plausible, given what Dennett says, to regard these as possessors of both causal relevance *and* causal power. First of all, illata figure in the (presumably causal) explanations of subpersonal psychology.¹⁰ Hence, they are possessors of causal relevance. Secondly, as physical states of organisms, states which are not abstract or calculation bound entities in the sense in which abstracta are, they will bestow upon objects which possess them dispositions to enter into causal relations; they have causal power. It is because intentional abstracta are realized by subpersonal illata which have causal power that intentional abstracta can have causal relevance.¹¹

9. "Three Kinds of Intentional Psychology", op. cit., pp. 55-56.

10. "Three Kinds of Intentional Psychology", op. cit.

11. "Three Kinds of Intentional Psychology", op. cit.

It has been argued that Dennett's illata/abstracta distinction can plausibly be seen as the distinction between those states which have both causal power and causal relevance (illata) and those states which merely have causal relevance (abstracta). I want to conclude this Appendix by pointing out some difficulties with Dennett's later position.

Dennett's later position emerged out of (i) a commitment to the constitutive role played by rationality in psychological ascription, and (ii) a commitment to viewing mental states as causally relevant entities. I will argue that Dennett's reconciliation of these two principles must be rejected. The problem is not so much with Dennett's distinction, but, rather, with the place he draws it.

It is the fact that individuation of mental states is essentially dependent on rationality assumptions that prevents the intentional states of folk psychology being illata: that is, being real causally powerful states of organisms. This status falls to illata. However, the states posited by folk psychology, conceived of as abstracta, can still be regarded as possessors of causal relevance because they supervene on, or are realized by, these underlying illata.

However, a puzzle now arises. If the fact that beliefs involve rationality prevents their being illata, why does the same fact not prevent their being abstracta. According to Dennett, it is facts about

an organism's possession of subpersonal illata which *make true* facts about that organism's possession of folk psychological abstracta. Put another way, Dennett holds something like the following supervenience thesis: if any two organisms are type-identical in point of their subpersonal illata, then they must be type-identical in point of their folk psychological abstracta. Now, given such a supervenience thesis, if subpersonal cognitive psychology is supposed to be literally true, and if subpersonal illata are states which are literally possessed by organisms, then folk psychological abstracta would have to be things literally possessed by organisms. For folk psychological states, conceived of as abstracta, are calculation bound entities, and facts of possession of these states by an organism are made true by facts of possession of their underlying illata. Then it seems that folk psychological states would have to be things literally possessed by certain systems. And this does not at all square with the fact that belief attributions essentially involve a rationality idealization that is only approximated but never realized. The crucial point is this: though the idealization view is perhaps harmless when intentional states are construed instrumentally, such a view is incoherent when such states are construed as abstracta whose possession by a system is realized by real underlying illata.¹²

12. This point is made by Paul Yu and Gary Fuller; "A Critique of Dennett", in Synthese, vol.86, no.3 (1986), pp. 453-76.

This is a serious problem for Dennett since, although he is prepared to be revisionary about many features of folk psychology, he is not prepared to abandon the view that intentional ascription presupposes ideal rationality. And, because of this problem, I believe that Dennett's distinction between abstracta and illata, his specific application of the distinction between causal power and causal relevance, must be rejected.

I have chosen to discuss Dennett's view because it is, I believe, symptomatic of a general tendency to view the explanations of folk psychology as explanations which go proxy for, and hence are to be replaced by, explanations of a mature cognitive science dealing with entities which are in some sense 'more real' than the entities of folk psychology. And this tendency, again, seems to stem from a specific application of Ockham's razor: The explanatory force of folk psychological states is realized by underlying states which are in some sense more real than those folk psychological states. Hence, to get at the 'real', scientifically utilizable, explanations we must get at the real underlying states, in this case Dennett's illata. And it is this form of the application of Ockham's razor, outlined in the first chapter, and represented here by Dennett which is the principle target of this thesis.

Chapter II:

Conceptual Explanatory Methodological Solipsism: Narrow Content, Expressibility, and Explanation.

1

In Chapter I it was argued that the principle of methodological solipsism could be given two, importantly distinct, interpretations. On the one hand, it might be interpreted as ontological methodological solipsism (OMS). On the other hand, it might be interpreted as explanatory methodological solipsism (EMS). The interpretation it receives in most current literature, indeed the interpretation it must receive if it is to do the work required of it, is of the explanatory variety.

Moreover, it was argued that within the EMS principle, a further distinction could and should be observed. This is the distinction between formal explanatory methodological solipsism (FEMS), and conceptual explanatory methodological solipsism (CEMS).

The subject of this chapter will be the CEMS principle. The remaining chapters will be taken up with discussion of FEMS. In many respects, the discussion of CEMS will be less original than the later treatment of FEMS. I choose to discuss the former first because there will emerge, in that discussion, several principles which together

constitute a *pattern*: a pattern that, it will be argued, also emerges in the discussion of FEMS.

In chapter I, the CEMS principle was defined as the claim that the causally relevant properties possessed by internal components (where the causal relevance, of course, is relevance of a sort appropriate to cognitive psychology) are properties of some form of content; they are the properties constitutive of what has become known as *narrow content*. This statement, though correct, does not bring out the force of the claim made by CEMS. I want to make a few subsidiary comments which will highlight this force.

Propositional attitudes have semantic content. Consider belief. The predicate '... believes that ...' seems to express a two place relation between a person and a sentence. Assume, for argument's sake, that this is so. The sentence falling within the scope of the '-that' is known as the belief's content sentence. It is this sentence which expresses the content of the belief. Explicating the content of the belief that *s*, then, comes down to explicating the meaning of the sentence *s*.

There is a fairly general consensus that the meaning of a sentence *s* is a function of the representational powers of *s*: the properties in virtue of which *s* represents a state of affairs in the world. This consensus has its expression in classical truth-conditions approaches to meaning. According to this type of approach:

(1) For s to mean that p is for s to be true to be true iff q , for some ' q ' having the same truth-conditions as ' p '.

This type of content, 'content₁', attaches to a belief in virtue of the referential truth-conditional properties of the belief's content sentence.

Some believe that a belief's content is exhausted by content₁.² However, another central pillar in the history of the theory of meaning is that meaning is a function of use: the meaning of sentence s is a function of the way s is used in language. Call this usage the *cognitive role* of s . Then:

(2) For s to mean that p is for s to have some cognitive role \emptyset such that ' p ' also has cognitive role \emptyset .³

1. Donald Davidson, "Truth and Meaning", Synthese, vol.17, (1967), pp. 304-23. Reprinted in Davidson Essays on Actions and Events, op. cit., pp. 17-36.

2. Davidson, "Truth and Meaning", *ibid.*, and especially "Empirical Content", in E. LePore ed., Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson (Oxford, Blackwell, 1985), pp. 320-32. Fodor, in Psychosemantics, thinks that content is almost exclusively constituted by referential truth-conditions. See ch.4. and ch.3, pp. 72-73.

3. A claim such as this would be foundational to any conceptual role semantics. See, Brian Loar, Mind and Meaning (Cambridge, Cambridge University Press, 1981). Hartry Field, "Logic, Meaning, and Conceptual Role", The Journal of Philosophy vol. LXXIV, pp. 379-410. Gilbert Harman, Thought, (Princeton, Princeton University Press, 1973). A conceptual role account of *part* of meaning is defended by McGinn, "The Structure of Content", op. cit. See also, N. Block, "Advertisement for a Semantics for Psychology", op. cit.

This type of content, content₂, attaches to a belief in virtue of the cognitive role of that belief's content sentence. There are various stories of how content can be constituted by functional role, but these are not of immediate concern here.⁴

Given the framework briefly sketched above, and given the assumption that part of a belief's content is constituted by the cognitive role of the belief, (an assumption granted by many, but by no means everyone), then we can identify what we can call the narrow semantic properties as those constituents of ordinary semantic content which are constituted by functional role, or, conversely, those constituents which are not constituted by the referential truth-conditional properties of the belief.

However, the defender of CEMS must claim far more than that part of a belief's content is constituted by cognitive role. He must claim that the narrow semantic properties thus realized have causal relevance: that they play an explanatory role in the causal explanations of cognitive psychology. Hence, statements of these explanations will contain terms or expressions which make direct reference to these narrow semantic properties. And a precondition of this is that these narrow semantic

4. The standard account runs as follows. There is an isomorphism between, on the one hand, logical relations holding between propositions or content-sentences, and, on the other, causal or functional relations holding between first-order internal states of organisms. According to this account, it is then possible to attribute content to these internal states on the condition that this isomorphism be preserved. The details are beyond the scope of this thesis. See Loar, Mind and Meaning, op. cit., and Fodor, Psychosemantics, ch.3, pp. 78-80.

properties be *abstractable* from ordinary wide semantic properties in such a way as to be capable of playing an explanatory role in such explanations.

Consider what would be meant by saying that ordinary wide semantic properties play an explanatory role in causal explanations. The most natural way of construing this is as the claim that statements of the causal explanations contain sentences which express ordinary wide semantic content. Similarly, then, the defender of CEMS, committed as he is to the causal relevance of narrow semantic properties, is also committed to the existence of sentences which express narrow content. And it will shortly be argued that there are no such sentences.

Put another way, I will from now on insist on a distinction between *narrow semantic properties* and *narrow content*. Narrow semantic properties will be understood as those components of semantic content which are constituted by cognitive role, where these properties cannot be abstracted from the semantic content of which they form a proper part. Narrow content, on the other hand, will be understood as a part of semantic content, constituted by cognitive role, which has been abstracted from that ordinary semantic content. Thus, in this sense, narrow content is a new form of content, a type of content which is narrow.

The point to emphasize is that it is possible to hold that there are narrow semantic properties without holding that there are narrow contents. One cannot infer from the fact that part of ordinary semantic

content is constituted by narrow semantic properties to the claim that there exists a distinct type of content - narrow content - capable of being abstracted from, and existing independently of ordinary wide semantic content. But this is just the move which the defender of CEMS must make: he is committed to the existence of a separate type of content, abstractable from ordinary content and, thereby capable of having causal relevance. And I shall argue that the notion of narrow content, understood in this sense, is literally incoherent.

2

In the immediate post-Putnam age, narrow content was often assimilated to *observational* or qualitative content. The motivation for this view clearly stems from Putnam's Twin Earth case. The moral of the Putnam story was that natural kind concepts are externally individuated.⁵ A corollary is that sentences such as 'water is wet', since they contained terms denoting natural kinds, could not be used to express any content which is genuinely narrow.

Nevertheless, it might be argued, my belief that water is wet and the corresponding belief of my counterpart are identical in point of cognitive or functional role, hence, identical in point of narrow content. And the reason that this is so is due to the fact that both water and retaw have associated with them the same

5. "The Meaning of 'Meaning'", op. cit.

qualitative/phenomenal/observational properties - they are both colourless, odourless, drinkable ... etc, etc. Hence, the next move seems *prima facie* plausible: narrow content can be specified in terms of observational or qualitative properties. For example, on this view it would be possible to specify the narrow content of the belief that water is wet in terms of a sentence something like, "The colourless, odourless, drinkable, transparent ... liquid is wet."⁶

The problem with this approach, however, turns on the nature of observational concepts - concepts such as *colourless, odourless, transparent ... etc* - in terms of which the above sorts of descriptions will be stated. For there are plausible considerations which suggest that these observation concepts are themselves externally individuated. The sources for this point are two.

The arguments of *Tyler Burge* were properly dealt with in chapter I. In that chapter it was pointed out that his arguments do not turn on the natural kind status of the external objects involved. In his example, arthritis is not a natural kind. However, Burge would go further and claim that his arguments can be applied to all concepts:

" Such arguments go through for observational and theoretical notions, for percepts as well as concepts, for natural kind and non-natural kind notions, for notions that are the special preserve of experts, and for what are known in the psychological literature as 'basic categories'. Indeed, I think that, at a minimum, relevantly similar arguments can be shown to go through with any notion that

6. An approach to narrow content based along something like these lines is provided by Dennett, "Beyond Belief", *op. cit.*

applies to public types of objects, properties, or events that are typically known by empirical means."⁷

Since observational concepts are apparently paradigm cases of notions which are "... known by empirical means", Burge would seem to be claiming that his arguments establish that observational concepts are externally individuated. However, when one goes on to apply Burge's arguments to observational concepts, one sees that the intuitions they evoke are far less clear cut than those implicated in his standard examples.

Consider Burge's earlier example, transposed to cover observational concepts. Oscar lives in a language community in the actual world. He has many beliefs, desires and other attitudes directed toward sentences of the form '... is red ...'. However, he also suffers from a misapprehension. He believes that the term 'red' applies not just to red things, but also to things which, in the rest of the community, are called yellow. So he believes things like daffodils are red etc.

Now the intuition which Burge must invite us to share in this case is that Oscar still has red-beliefs, red-desires, and other attitudes involving the property of redness. Burge must claim that the only deviancy in Oscar is that a lot of what he believes about the property

7. "Individualism and Psychology", op. cit., p. 6.

of redness is wrong. He still possesses beliefs directed toward the property of redness, it is just that many of his beliefs are wrong.

We now imagine a counterfactual situation. Counterfactually, Oscar lives in a language community in which the term 'red' applies not only to things which are red, but also to things which are yellow. The standard counterfactual use of the term 'red' encompasses Oscar's actual misuse.

The second intuition which Burge must invite us to share is that, in the counterfactual situation, Oscar does not have red attitudes. That is, Oscar counterfactually lacks the attitudes which we, in the actual world attribute with content clauses containing the term 'red' in oblique occurrence. Oscar in the counterfactual situation, unlike Oscar in the actual situation, does not have the concept of redness. He has the concept of counterfactual redness (= *reallow*, as we might say).

Burge's arguments are based upon intuitions. In the case of arthritis, it is intuition which tells us that Oscar has propositional attitudes directed toward propositions of the form '...arthritis...' in the actual, but not in the counterfactual, situation. And the problem is that in the case of observational concepts such as the concept of redness, the intuitions appear markedly less clear-cut than in the case of arthritis. The problem lies particularly with the first intuition. Would we really be willing to concede that, although Oscar thinks that the term 'red' applies both to red things and to yellow things, Oscar nonetheless has a concept of redness. The issues here are complicated,

and this is not the place to decide them. The relevant point is that Burge's arguments for the external individuation of observational concepts are not, without further argument grounding the intuitions they presuppose, anywhere near conclusive. Therefore, by themselves, they would constitute a fairly weak reason for rejecting the view that narrow content can be specified in terms of observational properties.

However, there is a far more simple, direct and powerful route to the conclusion that observational properties are externally individuated. Consider an individual who possesses certain mental states. The narrow content of those states is supposedly that form of content which can be specified without reference to anything outside the head of the individual.⁸ But specification of observational content does, quite clearly, make reference to external properties - it refers to observational properties, redness, wetness etc. Correlatively, the notion of external individuation is defined in terms of there being a necessary individuating relationship between contents and external properties. And this individuating relationship also appears to hold for observational contents: these contents are individuation dependent on the observational properties to which they refer. Therefore, observational contents are not genuinely narrow.

This point is so simple it is difficult to see why it has been overlooked. *Colin McGinn* has argued that the reason stems from a

8. This is the standard characterization of the notion of narrow content and stems from Putnam's "The Meaning of "Meaning".

tendency to conflate the question of the external individuation of a concept with the possibility of setting up, for that concept, a Twin Earth example of the *Putnam/Burge sort*. There has been a tendency to look at observational concepts, see that they are not susceptible to Twin Earth cases in the same way as, for example, natural kind concepts are, and conclude from this that they are not externally individuated. McGinn argues that this is a mistake, and derives from a failure to recognize an important distinction within the category of externalism.

The distinction is between weak and strong externalism.⁹ The statement of externalism as the claim that mental contents are individuated by the external objects and properties which they represent covers two, quite distinct, types of case:

(a) The objects and properties are causally (or, at least, environmentally) connected to the mental state whose content they individuate.

(b) The objects and properties need not be causally or environmentally connected to the mental state, but must occur in the same possible world as that mental state.

An interpretation of externalism along the lines of (a) yields strong externalism. An interpretation along the lines of (b) leads to weak externalism.

9. The distinction between strong and weak externalism derives from an unpublished lecture series given by McGinn, Oxford University, Hilary Term, 1987. Forthcoming in McGinn, Mental Content.

McGinn argues that strong externalism, while true of natural kind concepts, does not seem to be true of all concepts. In particular, it does not appear to be true of observational concepts. Consider how one might try to set up a Twin Earth thought experiment for observational concepts. In the case of natural kind concepts, such as water, the route is well known. The examples proceed by way of the following scheme:

	Earth	Twin Earth
Words	'Water'	'Water'
Experience	Colourless, odourless...	Colourless, odourless...
as of	... liquid	... liquid.
Properties	Water, i.e. H ₂ O	Retaw, i.e. XYZ.

However, consider now the corresponding case for qualitative concepts. Consider, for example the concept of redness:

	Earth	Twin Earth
Words	'Red'	'Red'
Experience as of	red	red
Properties	Red	Green

Consider the consequences that would obtain if the above scheme were cogent. The following case would be a possibility. Although I believe that I am having an experience of red, my experience is in fact an experience of green. The problem with this possibility, the reason that it cannot be taken to be a genuine metaphysical possibility, is that it unacceptably divorces qualitative experiences from beliefs about those experiences. And the reason this is a problem in the case of qualitative or observational concepts, but not in the case of natural kind concepts, is that qualitative concepts denote kinds which have no *real essence* which determines the extension, and hence, the correct application of those concepts independently of experience. The criteria of application of (terms denoting) qualitative concepts must be formulated in experiential terms; in terms of how they appear to the subject. And the above sorts of Twin Earth examples, if they could be coherently applied to qualitative concepts, would entail that experience was an inadequate basis both for the initial application, and for the repeated application of those concepts. But given that the kinds associated with qualitative concepts have no real essence, the only other basis there could be for classification and application of these concepts would be experiential in character. Therefore, if the above sorts of Twin Earth cases could be coherently applied to qualitative concepts, there would be no basis at all for application of these concepts. But this implies that we do not know the meaning of qualitative terms. And this is plainly unacceptable.

Considerations such as these indicate that qualitative or observational concepts are not individuated *strongly* externalistically.

And this gives further credence to the rejection of Burge's arguments concerning qualitative concepts. However, McGinn argues, this is quite compatible with adopting a *weak* externalism with regard to qualitative concepts. And, indeed, weak externalism does appear to be the correct position to adopt with regard to individuation of these concepts.

It is possible to set up a *modified* Twin Earth example to show why this is so. Oscar₁ lives in a world - the actual world - in which there exists the property of redness. So Oscar₁, on the basis of various patterns of experience can, correctly or incorrectly, attribute the property of redness to various objects. And Oscar₁ can have thoughts with contents expressed by sentences such as 'this shirt is red', 'this light is red', etc. Whether such thoughts are ultimately true or false, they are nonetheless thoughts about things which are *red*.

We now imagine a counterfactual situation. Counterfactual Oscar - Oscar₂ - lives in a world in which different evolutionary pressures have had the following result. In this world, there is no property of redness. The evolutionary concerns of Oscar₂'s species have not caused them to distinguish what we call red from what we call yellow. Instead, they run the properties together into an indivisible whole - the property of rellowness. What is crucial to this thought experiment is that members of the counterfactual linguistic group do not regard the property of rellowness as a composite property capable of being factored off into red and yellow components. Rather, the dominant evolutionary pressures have resulted in them regarding rellowness as a simple property, such as Oscar₁ regards redness.

Now suppose that throughout their lives Oscar₁ and Oscar₂'s non-intentionally described experiences have been identical. They have had the same streams of consciousness etc. The phenomenological characteristics of their experiences are, and always have been, type-identical. Still, it would not be true to say that they possess the same intentional states. Oscar₁'s beliefs are about *red* things and about *yellow* things. Oscar₂'s corresponding beliefs are about *reellow* things.

If either Burge's or McGinn's arguments are correct here, then observational concepts will receive some form of external individuation. And this means that there is a problem with explicating the notion of narrow content in terms of observational or qualitative concepts. For narrow content is supposed to be that independent and abstractable component of semantic content which respects methodological solipsism. Narrow content is supposed to be characterizable solely by reference to what occurs 'in the head' of a subject. Thus any characterization of narrow content which employs concepts which are externally individuated, even if the externality here is weak, will be incompatible with methodological solipsism. Therefore, the type of content arrived at via explication in terms of qualitative concepts will not be narrow at all. Narrow content cannot be characterized in terms of observational or qualitative concepts.

The problems associated with assimilating narrow content to qualitative content is indicative of a general problem in expressing narrow content with sentences of natural language. Natural language employs terms which express concepts which are *wide* in the sense of being (weakly or strongly) externally individuated. Hence, natural language cannot be used to express any version of content which respects methodological solipsism. One might seek to put the point thus: Narrow content is inexpressible in sentences of natural language. And perhaps this will indicate just how radically the notion of narrow content differs from the notion of ordinary content. For, according to standard interpretation, semantic content is just that which can be expressed in sentences. So, if one proposes to understand the notion of narrow content by analogy with ordinary content, then the notion of narrow content is *incoherent*.

The problem of the inexpressibility cum incoherence of narrow content is addressed in a recent book by Jerry Fodor.¹⁰ Corresponding to the notion of external individuation, Fodor talks of expressions of natural language being *anchored*, where, "... an expression is anchored iff it has a determinate semantic value".¹¹

10. Psychosemantics, op. cit., ch.2. Equivalently, "Individualism and Supervenience", op. cit.

11. Psychosemantics, op. cit., p. 49.

Fodor discusses Putnam's Twin Earth case. Consider the thoughts had by me and my Twin Earth counterpart, expressed by the sentence "Water is wet". These distinct thoughts are supposedly type-identical in point of narrow content. The question is, what is this content which the two thoughts have in common? Fodor argues that this question is ill advised since it admits of no answer. An answer would require the uttering of an English sentence which expresses just what my 'water'-thoughts have in common with those of my twin. And this cannot be done: " That's because the content that an English sentence expresses is ipso facto anchored content, hence, ipso facto, not narrow."¹² Of course, this is not a problem which affects just English. Twin-English, and, indeed, any language whose terms are 'anchored' will be equally affected. Fodor continues:

" So, in particular, qua expression of English 'water is wet' is anchored to the wetness of water (i.e. H₂O) just as, qua expression of Tw-English, 'water 2 is wet' is anchored to the wetness of water 2 (i.e. to the wetness of XYZ). And, of course, since it is anchored to water, 'water is wet' doesn't - can't - express the narrow content that my water-thoughts share with my Twin's Narrow content is radically inexpressible because it is only content potentially; it's what gets to be content when - and only when - it gets to be anchored."¹³

Fodor, here, describes essentially the same position as that advanced at the beginning of this section. Narrow content cannot be expressed in sentences of natural language. Hence, narrow content is not the same

12. Psychosemantics, op. cit., p. 50..

13. Psychosemantics, op. cit., p. 50.

sort of thing as ordinary semantic content at all. It is only content *potentially*, as Fodor puts it.

Consider the consequences this has for CEMS. According to this principle, it is the narrow content of mental states which has causal relevance. That is, according to the above principle, it is the narrow content of mental states which will play an explanatory role in the causal explanations of cognitive psychology. Therefore, statements of these explanations will have to contain sentences which express the narrow contents of mental states. But, if the previous arguments are correct, the idea of a sentence which expresses narrow content is literally incoherent. Sentences (of all natural languages) employ terms which express concepts which are individuated widely. Therefore, there just are no sentences which could be used by a cognitive psychology which respects CEMS. Therefore, cognitive psychology cannot possibly respect CEMS. This completes the argument *against* the principle of conceptual explanatory methodological solipsism.

4

If narrow content is understood as an independent and abstractable form of semantic content, then the notion of narrow content is incoherent. The only grip we have on the notion of narrow content is by way of the ordinary wide semantic content of which it forms a proper (though dependent) part. It is this fact which provides the basis of our talk of narrow semantic properties where these are understood as

non-abstractable dependent parts of ordinary semantic content. And, it is this fact, it will be argued, which enables us to see why ordinary wide semantic content can play an explanatory role in the explanations of psychology.

As we have seen, the inexpressibility of narrow content is conceded by Fodor. What, then, allows us to coherently speak of mental states possessing, or being identical in point of, narrow content. Fodor's answer is that narrow content is *mentioned* by ordinary wide sentences of natural language:

" The corresponding consideration is relatively transparent in the case of demonstratives. Suppose the thought 'I've a sore toe' runs through your head and also runs through mine; what's the content that these thoughts share? Answer: you can't say what it is by *using* a sentence since, whenever you use a sentence that contains 'I', the 'I' that it contains automatically gets anchored to you. You can, however, sneak up on a shared content by *mentioning* a sentence, as I just did above. In such cases, mentioning a sentence is a way of abstracting a form of words from the consequences of being anchored....

.... Referring expressions of English can therefore be used to pick out narrow contents via their hypothetical semantic properties. So, for example, the English expression: 'the thought that water is wet' can be used to specify the content of a mental state that my twin and I share (even though, qua anchored to H₂O it doesn't, of course, *express* that content).... Roughly speaking, this tactic works because the narrow thought that water is wet is the *unique* narrow thought that yields the truth-condition H₂O is wet when anchored to my context, and the truth-condition XYZ is wet when anchored to his."¹⁴

This account indicates why we are justified in talking of narrow semantic content at all. This idea is the basis of the extensional

14. Psychosemantics, op. cit., pp. 50-51.

criterion for narrow content identity provided by Fodor.¹⁵ Moreover, the distinction between expressing and mentioning provides the beginnings of an account of why it is possible to regard semantic contents as bearers of causal relevance. Fodor's view is that sentences which express ordinary semantic content can, thereby, *mention* narrow content. I think Fodor's idea can be usefully developed in terms of the notion of *indexing* outlined in chapter I. It is not being claimed, of course, that Fodor would endorse the following account of indexing. It is worth pointing out, however, that Fodor is in need of some distinction between causal power and causal relevance. For he has at the same time committed himself to the view that the explanations of cognitive psychology are to be based around the notion of narrow content, *and* to the view that there are no sentences which express narrow content.¹⁶

The use of indexing in the philosophy of mind exhibits close similarities to its use in the theory of measurement. The idea is perhaps most completely worked out by Brian Loar in Mind and Meaning.¹⁷ The basis of Loar's idea is that the relations between a person's physical properties are somehow structurally isomorphic to some systematic framework that can be specified by mentioning propositions,

15. Psychosemantics, pp. 47-48.

16. Fodor ends up in this predicament because he fails to distinguish adequately causal power and causal relevance. This is a problem which seems to affect much of Fodor's work. The failure is reflected in a tendency to vacillate between strong and weak versions of the representational theory of the mind. This point will be discussed in Chapter V.

17. Loar, Mind and Meaning, op. cit.

their logical interconnections, etc.¹⁸ That is, there is an isomorphism of logical relations holding between propositions and causal or functional relations holding between first-order physical states. In virtue of this, a proposition can index the functional role possessed by a given physical state:

" ... certain logical relations among propositions, and correspondence relations between propositions and the world, mirror counterfactual relations among certain first-order states, and between some of these and extra-mental input conditions. So there is an isomorphism between the two networks. By virtue of its unique place in the logical network (which remains to be specified) a proposition can index a belief's unique functional role. It must be emphasized that the proposition is extrinsic to the underlying state."¹⁹

The notion of indexing can be given application to show why it is possible to regard ordinary semantic contents as bearers of causal relevance.

Recall the case of numbers. Properties of number, it was argued have causal relevance even though they are not possessors of causal power. The reason is that numbers are indices of intrinsic properties of objects, and it is these intrinsic properties which are possessors of causal power.

We can take a similar line with semantic contents. Semantic contents do not have causal power. However, it can be argued, they do

18. Mind and Meaning, op. cit.

19. Mind and Meaning, op. cit. Field gives a similar account in the appendix to "Mental Representation", op. cit.

have causal relevance, and the reasons for this are largely the same as in the case of numbers. Firstly, relations between propositions which express semantic content are isomorphic with relations with relations holding between certain, presumably physical, states of organisms. Because of this isomorphism, the propositions can index their corresponding physical states. It is these physical states which are possessors of causal power. And, then, semantic contents can have causal relevance because the propositions which express them index intrinsic causally powerful states of organisms; just as properties of number can have causal relevance because numbers index causally powerful states of objects.

This, then, is the story. Semantic contents can play an explanatory role in the causal explanations of psychology because the propositions which express them index intrinsic causally powerful states of organisms. A few points are needed to tidy up this matter.

Firstly, if one does not like talk of propositions, this can easily be replaced with talk of *sentences* which express content. The intrinsic causally powerful states will in this case be indexed by sentences, and semantic contents will possess causal relevance in virtue of being expressed by these sentences.

Secondly, the above account tries to show how semantic contents can figure in causal explanations independently of possessing causal power. But this is just one possible explanation of a general phenomenon, viz. the distinction between causal power and causal relevance.

Consequently, one is free to maintain the distinction even if one does not like the particular means of its implementation I have described here.

Thirdly, it should be emphasized that the propositions/content sentences are *extrinsic* to the physical states they index. Therefore, although sentences have logical structure, there is no need to suppose that the states which they index must have logical structure. Now, given that these physical states are what has become known as *internal representations* - a subject treated in much more depth later - there is nothing in the indexing model which supports the language of thought hypothesis that internal representations are sentence-like entities which have logical structure. On the other hand, there is nothing here which is incompatible with that hypothesis either.²⁰

Finally, a few notes on the notion of *isomorphism* are in order. This is an extremely vague notion, incorporated in the present discussion only because of the role it plays in other discussions of indexing.²¹ This being said, the notion cannot be taken too literally:

20. Thus Field is wrong in his claim that the indexing model entails that mental representations have logical structure. See the appendix to "Mental Representation", *op. cit.* This point was first made in Chapter I, n.11.

21. This point will be further discussed in the Conclusion. There it will be recommended that the notion of isomorphism be dropped altogether from the model of indexing.

Firstly, nothing has been said which entails that the isomorphism must be total. Rather, it may be that indexing can proceed on the basis of the establishing of *partial* isomorphisms between sets of sentences or propositions and internal representations.

Secondly, the mapping relation on which the isomorphism is based need not be a simple one to one relation but may be extremely *complex*, consisting of many-one and one-many relations.

Thirdly, and most importantly, the notion of isomorphism at work here is a *normative* one. The logical relations into which sentences or propositions enter with one another play a normative role with respect to the physical states or internal representations which they index. That is, the relations entered in to by propositions play the role of an *ideal* which relations between physical states can mirror or correspond to only imperfectly. This topic will be discussed at length later in the thesis.

In short, the basis of the claim made here is that the distinction between causal power and causal relevance, together with a (non-mandatory) story of its implementation based around the notion of indexing, shows that the constraint placed on psychological theorizing by CEMS is misplaced. Properties of ordinary semantic content can have causal relevance - can play an explanatory role in the causal explanations of cognitive psychology - even if they do not have causal power.

At the beginning of the chapter I alluded to a *pattern*. The pattern, it can now be seen, is this:

1. The only grip we have on the notion of narrow content is provided by our knowledge of propositions or sentences which express ordinary wide semantic content. That is, narrow content cannot be identified independently of wide semantic content. Therefore, in the formation of psychological theories, CEMS cannot be complied with.
2. The constraint expressed by CEMS is, anyway, misplaced. Ordinary semantic content can have causal relevance even if it does not have causal power, or is not taxonomized according to causal power.

In the remaining chapters, I will be discussing the principle of formal explanatory methodological solipsism. It will be seen that an identical pattern emerges there. Firstly, due to problems of independent identifiability, FEMS cannot be complied with. Secondly, due to the distinction between causal relevance and causal power, the demand for FEMS is misguided anyway.

Longer Note 2.1. In Chapter II a certain strategy for deriving narrow content from ordinary wide content was described and criticized. We might call this the "Phenomenological Strategy". (The expression is due to LePore and Loewer, "Solipsistic Semantics", Midwest Studies in Philosophy, vol. X (1986), pp. 595-614, pp. 606-608. It consisted in finding a collection of concepts which are not externally individuated and constructing all other concepts out of these. The approach can be labelled "phenomenological" since the envisaged base solipsistic concepts were qualitative, observational, or phenomenological in character. It was argued that this strategy fails because even phenomenological concepts are externally individuated. They do not have solipsistic semantics.

There is, however, another strategy present in the literature which one might call the "Indexicalist Strategy". (The expression is again due to LePore and Loewer, *op. cit.*). This involves treating belief content indexically in a way that is supposed to presuppose the existence of no entities external to the thinker. This strategy is associated with John Searle, Intentionality (Cambridge, Cambridge University Press, 1983). For example, Searle says that "water" is to be defined indexically as whatever is identical in structure with the stuff causing *this* visual experience. However, such an approach suffers from closely related difficulties to those of the phenomenological strategy. For a start, it is clear that the notion of 'stuff' is externally individuated. Consider also the notion of causation. It seems possible to set up a Twin Earth case for this concept. Suppose that in this world, Hume (or the standard interpretation of Hume) is correct. So, when I have a belief attributable by a sentence of the form 'Causation is universal', among the external properties I will be related to will be properties of contiguity, succession and constant conjunction. But I will *not* be related to any *external* property of natural necessity. However, my Twin lives in a world where Hume is wrong, and (what he calls) 'causation' is a matter of natural necessity. So when he has a belief attributable by a sentence of the form 'Causation is universal', as well as being related to external properties of contiguity, succession and constant conjunction, he will also be related to the external property of natural necessity. Hence, he has a different thought. The notion of causation, then, seems to be *strongly* externally individuated. Another problem with Searle's indexicalist strategy is pointed out by Ernest LePore and Barry Loewer, "Solipsistic Semantics", *op. cit.*, p. 608. The expression 'causing this visual experience' does not uniquely refer. "There are many things causing this visual experience. One of the causes is the pattern of neuron firing in the optic nerves. We could exclude this by adding 'the external liquid stuff causing ...' but now we are faced with the problem that 'external liquid' does not have SS [solipsistic semantics]."

Fodor's extensional criterion of identity for narrow content derives from the indexicalist strategy. See Fodor, Psychosemantics, *op. cit.* And his account originated with Stephen White, "Partial Character and the Language of Thought", Pacific Philosophical Quarterly, vol. 63 (1982), pp. 37-65. The criterion is based on the following. Consider the entire environment of myself or my Twin as a context. Then my belief that 'water is wet' can be interpreted as expressing a character

(in Kaplan's sense) that maps my context onto the content that H₂O is wet, and my Twin's corresponding belief maps his context onto the content that XYZ is wet. This procedure gives one a criterion of identity for a content which is narrow. But it cannot serve as an expression of narrow content. As LePore and Loewer point out, it yields no specification of content that can follow "believes that". It is one thing to have a criterion of identity for thoughts, it is quite another to be able to say what the content of a thought is. But as LePore and Loewer again point out, if narrow contents are to figure in the explanations of psychology we need to have more than simply a criterion of identity for such contents, we need to be able to say what those contents are.

The Anomalism of Mental Properties

1

The principle of the anomalism of mental properties is a principle largely associated with the work of Donald Davidson.¹ Davidson states the principle as follows:

".... there are no strict deterministic laws on the basis of which mental events can be predicted and explained (the Anomalism of the Mental)."²

There are two points to note about the above formulation. Firstly, the word 'deterministic' does no useful work here. Indeed, as shall be seen shortly, its inclusion is inconsistent with what Davidson says elsewhere.³ Hence, in all future reference to the principle of anomalism, the condition of determinism should be understood as omitted. Secondly, Davidson is not claiming simply that as a matter of fact there

1. Donald Davidson: "Mental Events", in L. Foster and J.W. Swanson eds., Experience and Theory, (University of Massachusetts Press, 1970), pp. 79-101. Reprinted in Davidson, Essays on Actions and Events, op. cit., pp. 207-25. "The Material Mind", in P. Suppes ed., Logic, Methodology and the Philosophy of Science, vol.4 (North Holland, Amsterdam, 1973). Reprinted in Essays on Actions and Events, op. cit., pp. 245-59. "Psychology as Philosophy", in S.C. Brown ed., Philosophy of Psychology (London, MacMillan Press, 1974). Reprinted in Essays on Actions and Events, op. cit., pp. 229-39.

2. "Mental Events", p. 208.

3. Specifically, with his claim that the ultimate comprehensive closed theory might be probabilistic. *Vide infra*.

are no laws on the basis of which mental events can be predicted and explained. His claim should not be interpreted as a de facto claim about the way the world happens to be. Rather, his claim is that there *cannot* be such laws. It will become clear that the force of the necessity here is conceptual or logical: there is something about the nature of mental concepts which makes laws involving mental predicates or properties a conceptual impossibility. Hence, a more accurate statement of the principle of anomalism would be:

Anomalism of the Mental: There cannot be strict laws on the basis of which mental events can be predicted and explained.

Any law which subsumes mental events has to be either a psychophysical or psychological law. Therefore, this claim can usefully be split into two subsidiary claims, which I shall call:

(i) *The Principle of Psychophysical Anomalism*: there cannot be strict psychophysical laws on the basis of which mental events can be predicted and explained.

(ii) *The Principle of Psychological Anomalism*: there cannot be strict psychological laws on the basis of which mental events can be predicted and explained.

Together, (i) and (ii) constitute the Principle of the Anomalism of the Mental. Proper understanding of the principle, then, involves understanding Davidson's notions of (a) the mental, (b) laws, and (c)

strictness. I have nothing new to say on these matters. The following comments are included for completeness.

1. *The scope of the mental.* Davidson proposes the following criterion for counting an event as mental:

" ... an event is mental if and only if it has a mental description, or (the description operator not being primitive) if there is a mental open sentence true of that event alone."⁴

A description, or open sentence, is mental if and only if it contains an essential occurrence of a mental verb.⁵ (The occurrence must be essential to rule out the possibility of there being a mental description or open sentence that is logically equivalent to a description or open sentence that does not contain a mental predicate).⁶

A verb is mental if and only if it expresses a propositional attitude. Davidson characterizes the notion of a propositional attitude by enumeration; propositional attitudes are such things as believing, desiring, intending, deciding, realizing, hoping, wishing, fearing, and the like.⁷ It is characteristic of verbs of propositional attitudes that their occurrence in a predicate creates a non-extensional context.⁸

4. "Mental Events", op. cit., p. 211.

5. "Mental Events", op. cit., p. 211.

6. "Mental Events", op. cit., p. 211.

7. "Mental Events", op. cit., p. 210.

8. "Mental Events", op. cit., p. 210.

It has often been thought that Davidson's criterion is too narrow. Because it is specified in terms of a verb that expresses a propositional attitude, it does not seem to apply to sensations, and these are traditionally regarded as paradigmatically mental.⁹ However, in fact, the criterion suffers from the converse problem: it is too broad rather than too narrow. The reason is that any event can be given a mental description, and so count as mental by the criterion.¹⁰

Davidson acknowledges this.¹¹ However, he says that this creates no problem given his purpose of showing that every mental event is a physical event. Problems would arise, he claims, only if the criterion failed to apply to some bona fide mental event.¹² This response, however, does not square at all well with Davidson's later argument for token identity. For this argument depends on the claim that mental events causally interact with non-mental events.¹³ And this would seem to require a more restricted notion of mental event.

2. *The Notion of Law.* Davidson characterizes laws as true lawlike sentences. Within this characterization, however, laws are not

9. This objection is raised by J.J.C. Smart in his "Davidson's Minimal Materialism", in B. Vermazen and M. Hintikka eds., Essays on Davidson: Actions and Events (Oxford, Oxford University Press, 1985).

10. This point is made by Brian McLaughlin, "Anomalous Monism and the Irreducibility of the Mental", in E. LePore and B. McLaughlin eds., Actions and Events: Perspectives on the Philosophy of Donald Davidson (Oxford, Blackwell, 1985), pp. 331-68, pp. 336-37.

11. "Mental Events", op. cit. pp. 211-12.

12. "Mental Events", op. cit., p. 212.

13. "Mental Events", op. cit., pp. 208 & 224.

delineated by form, but by function. Laws can be universalized conditionals, universalized biconditionals, or statistical in form. What is characteristic of them is that (i) if true they support counterfactuals, and (ii) they are confirmable by their positive instances.¹⁴

By 'a strict psychophysical law' Davidson means a strict law that features both mental and physical predicates; by 'a strict psychological law', a law that features only mental predicates.

The view of laws as sentences, adopted by Davidson, has been challenged.¹⁵ Some maintain that laws are relations between certain sorts of universals or 'natural properties'.¹⁶ The concerns of this thesis are largely independent of whether laws are sentences or relations between universals. However, later in this chapter it will be seen that the supervenience relation can be more simply and adequately discussed if one casts the discussion in terms of properties rather than predicates.¹⁷ Accordingly, the discussion of psychophysical laws will

14. "Mental Events", op. cit. p. 217.

15. D. Armstrong: What is a Law of Nature? (Oxford, Oxford University Press, 1983). M. Tooley "The Nature of Natural Laws", in Canadian Journal of Philosophy, vol. VII, (1977), pp. 667-98. David Lewis, "New Work for a Theory of Universals", Australasian Journal of Philosophy, vol.61, no.4, (1983), pp. 343-77. Chris Swoyer, "The Nature of Natural Laws", Australasian Journal of Philosophy, vol.60, no.3, (1982), pp. 203-23.

16. The notion of a natural property is due to David Lewis, *ibid.*

17. The crucial difference here is that properties can be combined together by way of infinite construction operations - for example, conjunction and disjunction - in such a way as to form new properties. The corresponding claim does not hold true for predicates. One cannot relate predicates under, for example, infinite disjunction in such a way as to form a new predicate.

be recast in terms of properties rather than predicates. Laws will be regarded as relations between properties, and the corresponding sentences regarded as statements of those laws. Nothing crucial to Davidson's arguments turns on this switch.

3. *Strictness.* I have nothing new to say on Davidson's notion of completeness, and my treatment of this complex subject will be brief in the extreme.

Firstly, the distinction between strict and non-strict laws is identical with the distinction between what Davidson calls homonomic and heteronomic laws:¹⁸

" On the one hand there are generalizations whose positive instances give us reason to believe the generalization itself could be improved upon by adding further provisos and conditions stated in the same general vocabulary. Such a generalization points to the form and vocabulary of the finished law: we may say that it is a *homonomic* generalization. On the other hand there are generalizations which when instantiated may give us reason to believe there is a precise law at work, but one that can be stated only by shifting to a different vocabulary. We may call such a generalization heteronomic."¹⁹

Homonomic laws are strict, heteronomic laws are non-strict.

Davidson develops the idea of homonomicity in terms of the notion of a *closed comprehensive theory*.²⁰ The characterization of the notions

18. "Mental Events", op. cit., pp. 219-20.

19. "Mental Events, op. cit., p. 219.

20. "Mental Events, op. cit., p. 219.

of closedness and comprehensiveness I will borrow from an article by W.S. Stanton.²¹

Following Stanton, let us introduce the notion of a *T-description* of an event. Given a theory T, a T-description of an event is a description couched solely in terms of the vocabulary of T. More precisely, an event is T-described if and only if there is some description under which the event is subsumed under or entailed by a law of theory T.²²

Stated in terms of this apparatus, Davidson's notion of closure is definable as follows: a theory T is *closed* if and only if events within the domain of T causally interact only with other events within T's domain. That is, T is closed if and only if there is no T-described event which causally interacts with an event which is not T-described.²³

Although Davidson does not explicitly say what comprehensiveness is, the required notion can be rendered fairly precise using the apparatus of T-descriptions: a theory T is *comprehensive* if and only if T is closed and every T-event uniquely satisfies a T-description under which it instantiates a law of T from which it can be predicted more accurately than by a law of any theory other than T.²⁴

21. W.S. Stanton, "Supervenience and Psychophysical Law in Anomalous Monism", Pacific Philosophical Quarterly, vol.64, no.1, pp. 72-79, especially p. 74.

22. "Supervenience and Psychophysical Law in Anomalous Monism", op. cit., p. 74.

23. "Supervenience and Psychophysical Law in Anomalous Monism", op. cit., p. 74.

The notion of strictness can now be given the following characterization:

A law is strict if and only if it exclusively features properties which are either, (a) properties of a closed comprehensive theory, or (b) properties constructed out of those of a closed comprehensive theory, or (c) properties reducible by bridge law to those of a closed comprehensive theory.²⁵

2

Davidson's arguments for the principle of the anomalism of the mental can be divided into two. These are the argument for psychophysical anomalism and the argument for psychological anomalism. In neither case does Davidson argue that there can be no laws involving mental properties, he claims only that such laws cannot be *strict*. For example, he allows that there may be laws of psychology which are more precise than laws of many other sciences - for example, meteorology and geology.²⁶ Davidson only denies that there can be *strict* laws of psychology.

24. "Supervenience and Psychophysical Law in Anomalous Monism", op. cit., p.74.

25. c.f. McLaughlin, "Anomalous Monism and the Irreducibility of the Mental", op. cit., p. 345.

26. Davidson: "Comments and Replies: Reply to Peters", in Essays on Actions and Events, op. cit., pp. 239-44, p. 240

1. *Psychological Anomalism*. The central line of argument runs as follows.²⁷

(1) Mental properties are not properties invoked, or quantified over, by a closed comprehensive theory.

(2) Mental properties cannot be linked via bridge laws with properties of a closed comprehensive theory.

∴ (3) There can be no strict psychological laws.

Davidson regards (1) as unproblematic. And (3) follows from (1) and (2) given the earlier characterization of strictness. The bulk of Davidson's further argument goes towards defending premise (2) - the denial of the possibility of bridge laws linking psychological properties with the properties of a closed comprehensive theory.

2. *Psychophysical Anomalism*. The argument against strict psychophysical laws is basically identical to the one given above with the qualification that the closed comprehensive theory in question is a physical theory.²⁸

(4) Mental properties are not properties invoked, or quantified over, by a closed comprehensive *physical* theory.

27. c.f. McLaughlin, "Anomalous Monism and the Irreducibility of the Mental", op. cit., p. 349.

28. c.f. McLaughlin, "Anomalous Monism and The Irreducibility of the Mental", op. cit, pp. 349-50.

(5) Mental Properties cannot be linked by bridge laws with properties of a closed comprehensive *physical* theory.

∴ (6) There can be no strict Psychophysical Laws.

The arguments are identical except for the above mentioned qualification. It is evident that Davidson takes the argument against strict psychophysical laws to be, in some sense, the more basic of the two. For example, he says, "... this [the fact that there cannot be strict psychophysical laws] is not quite the principle of the anomalism of the mental, but on reasonable assumptions entails it."²⁹ The "reasonable assumption", then, is the assumption that any closed comprehensive theory is a physical theory. Davidson is explicit in making this assumption. Hence, the argument (1)-(3) collapses into the argument (4)-(6).

This being said, the crucial premise is (5): the claim that mental properties cannot be linked by bridge laws with properties of a closed comprehensive physical theory. It is the alleged impossibility of psychophysical bridge laws which is used to establish the impossibility of any other sorts of law which feature mental properties. Accordingly, it is to this claim that Davidson directs the bulk of his argument.

29. "Mental Events", op. cit., p. 209.3

The basis of Davidson's argument against the possibility of psychophysical bridge laws - laws correlating mental properties with the properties of a closed comprehensive physical theory - is the claim that there is some sort of radical difference between the mental and the physical that makes such laws impossible.

" ... to allow the possibility of such laws would amount to changing the subject. By changing the subject I mean here: deciding not to accept the criterion of the mental in terms of the vocabulary of the propositional attitudes."³⁰

" There are no strict psychophysical laws because of the disparate commitments of the mental and physical schemes. It is a feature of physical reality that physical change can be explained by laws that connect it with other changes and conditions physically described. It is a feature of the mental that the attribution of mental phenomena must be responsible to the background of reasons, beliefs, and intentions of the individual. There cannot be tight connections between the realms if each is to retain allegiance to its proper source of evidence."³¹

" Standing ready, as we must, to adjust psychological terms to one set of standards and physical terms to another, we know that we cannot insist on a sharp and lawlike connection between them."³²

As a start at the interpretation of Davidson's arguments against psychophysical bridge laws it is important to clearly distinguish: *predicates, properties, and concepts*. A predicate is a term or

30. "Mental Events", op. cit., p. 216.

31. "Mental Events", op. cit., p. 222.

32. "Psychology as Philosophy", op. cit., p.239.

expression of language which denotes a property and expresses a concept. In virtue of this relation, concepts provide what I shall call the *criteria of application* of predicates. In virtue of this relation also, concepts provide what I shall call the *conditions of attribution* of properties. That is, the conditions under which we would attribute a property P receive specific expression in the concept associated with property P. Davidson's arguments against psychophysical bridge laws often tend to collapse the distinction between predicate and corresponding property. However, I will discuss the arguments in terms of concepts and *properties*.

The basic idea, then, of Davidson's arguments against psychophysical bridge laws is that there is a certain lack of fit between mental and physical *concepts*, and this lack of fit precludes mental and physical *properties* figuring in laws together. Davidson develops this idea by way of the introduction of the notion of a *constitutive principle*. The argument can be represented thus:

- (i) Attribution of mental and physical properties to organisms are governed by distinct sets of constitutive principles.
- (ii) These two sets of principles are disparate, essentially different.
- (iii) This difference makes strict psychophysical bridge laws impossible.

Attribution of a property is determined by the concept expressed by that property's corresponding predicate. So, constitutive principles

can be understood as principles describing essential interrelations between concepts. And then, the idea is that there cannot be strict psychophysical bridge laws because, " ... the constitutive principles of the mental are different from the constitutive principles of the physical in such a way that concepts from the two schemes cannot be joined so as to constitutively underwrite such laws."³³

According to Davidson, constitutive principles are synthetic a priori generalizations.³⁴ For a principle to have constitutive status with regard to the attribution of a certain range of properties is for it to play an essential role in the attribution of those properties. And constitutive principles have that status because they partly determine what counts as correct or incorrect attribution of a property. It is concepts which provide conditions of attribution of properties, and constitutive principles, by describing *regulative* and thereby *normative* relations between concepts, play a regulative and thereby normative role in the attribution of properties. Constitutive principles partly determine what can and cannot count as a reason for believing that a property can, in any particular case, be correctly attributed. The principles have an a priori status relative to our attribution of properties in particular cases. That is, our commitment to these principles cannot be jeopardized or undermined by any number of attributions of properties in particular cases. On the contrary, our

33. McLaughlin, "Anomalous Monism and the Irreducibility of the Mental", op. cit., p.353.

34. "Mental Events", op. cit., p. 221.

attribution of properties to particular cases is judged by its conformance to these principles.

The examples of constitutive principles of the *physical* listed by Davidson include ones concerning the measurement of length, mass, time, and temperature. These provide the "simplest possible" illustrations of constitutive principles of the physical³⁵). For example, the law of transitivity, (L) $L(x,y)$ and $L(y,z) \rightarrow L(x,z)$, is constitutive of the concept of length in that, unless this law holds, it is not possible to make sense of the concept of length. The constitutive principles of the physical will thus include ones concerning the transitivity, and also ones concerning other relations, such as the asymmetry, of the two-place relations longer than, heavier than, warmer than, later than, and so on. And it includes other principles that serve to distinguish the relations in question.³⁶

The constitutive principles of the mental, according to Davidson, include *principles of rationality*: He writes:

" Just as the satisfaction of the conditions for measuring length or mass may be viewed as constitutive of the range of application of those sciences that employ these measures, so the satisfaction of conditions of consistency and rational coherence may be viewed as constitutive of the range of application of such concepts as those of belief, desire, intention and action."³⁷

35. "Mental Events", op. cit., p. 220.

36. "Mental Events", op. cit., p. 220.

37. "Psychology as Philosophy", op. cit., pp. 236-37.

The role of the constraints of consistency and rational coherence in the attribution of mental properties is one of the concerns which dominates Davidson's work. A proper examination of this topic would exceed the means of this thesis, although a later chapter will touch on this subject. The key idea, however, is that some degree of rationality is implicated in all attributions of mental properties. Interpretation of the behaviour of an individual proceeds on the basis of attributing to that individual, firstly, actions which are characterized intentionally, and secondly, intentional states which together stand to those actions in relations of consistency and rational coherence. Thus, the intentional states are, together, said to *rationalize* the actions. That is, the intentional states, in virtue of their propositional content, provide the individual's rationale for producing the intentional action in question. And this procedure cannot be accomplished without at least a minimal degree of rational coherence.

According to Davidson, the constitutive principles governing the mental and those governing the physical are disparate, essentially different. The difference seems to be simply that the constitutive principles of the physical do not include principles of rationality. In contrast, the principles constitutive of the mental say what intentional states an individual ought rationally to be in, given that he is also in

38. These issues are addressed in Davidson, "Radical Interpretation", Dialectica, vol. 27, (1973), pp. 313-28. Reprinted in Inquiries into Truth and Interpretation, op. cit., pp. 125-39. Also, "Belief and the Basis of Meaning", Synthese, vol. 27, (1974), pp. 309-23. Reprinted in Inquiries into Truth and Interpretation, op. cit., pp. 141-54.

certain other states. They do this because they specify normative relations between propositional mental states that hold in virtue of the propositional contents of those states. In contrast, although physical constitutive principles play a normative role in the attribution of physical properties, the norms they invoke are not norms of rationality.

It is precisely because the constitutive principles of the mental include principles of rationality - i.e. principles which express norms of rationality - that there can be no strict psychophysical bridge laws.

" Any effort at increasing the accuracy and power of a theory of behaviour forces us to bring more and more of the whole system of the agent's beliefs and motives directly into account. But in inferring this system from the evidence, we necessarily impose conditions of coherence, rationality and consistency. These conditions have no echo in physical theory, which is why we can look for no more than rough correlations between psychological and physical phenomena."³⁹

Why does this difference in constitutive principles make psychophysical bridge laws impossible? Davidson is, to say the least, less than clear on this point. Probably the best interpretation and explanation of Davidson's argument here is given by Jaegwon Kim, in his recently published "Psychophysical Laws".⁴⁰ It is to this interpretation that I will adhere in this section.

39. "Psychology as Philosophy" *ibid.*, op. cit., p. 231.

40. Jaegwon Kim: "Psychophysical Laws", in LePore and McLaughlin eds., Actions and Events: Perspectives on the Philosophy of Donald Davidson, op. cit., pp. 369-86.

It has been argued that concepts can be viewed as providing conditions of attribution for properties. Now let us say that a property has *rational conditions of attribution* if and only if its attribution is in part governed by principles of rationality. And the attribution of a property will be governed by principles of rationality if and only if the concept associated with the property is a concept whose relations to other concepts can only be described in terms of rationality. A property whose attribution is not governed even in part by principles of rationality has *nonrational conditions of attribution*. Mental properties, in virtue of their associated concepts, have rational conditions of attribution. Physical properties, in virtue of their associated concepts, have nonrational conditions of attribution.

Now I shall avail myself of an idea put forward by Kim.⁴¹ The idea is this: bridge laws *transmit* criteria of application from the property referred to by the reducing predicate to the property referred to by the reduced predicate. So if X provides the conditions of attribution of the property referred to by the reducing predicate in a bridge law, and A is the property referred to by the reduced predicate in that law, then X is a condition of attribution of A. Let us apply this to the case of psychophysical bridge laws.

Let m_1 be a mental predicate, and let n_1 be a physical predicate. Let C_1 be a criterion of application of n_1 . Articulation of such a

41. "Psychophysical Laws", op. cit., especially pp. 374-81.

condition is probably a complex affair, but what is crucial is that this articulation will be regulated by the constitutive rules and principles of physical theory, not by those of the mental.

Kim points out, and we shall see that this claim is crucial, that if C_1 is a condition of attribution of n_1 , the connection between C_1 and n_1 must be more than a mere de facto concomitance; the connection must have modal force. There must be a necessary connection (for some form of necessity) between C_1 and n_1 .

Given these two conditions, Kim represents Davidson's argument as follows:⁴²

$$(1) \quad \Box (x)[C_1x \rightarrow n_1x]$$

and suppose we have the psychophysical law,

$$(2) \quad \Box (x)[n_1x \leftrightarrow m_1x]$$

whence;

$$(3) \quad \Box (x)[C_1x \rightarrow m_1x]$$

It is the conditional expressed by (3) which is incompatible with Davidson's claims regarding the categorical difference between mental and physical. The mental property m_1 receives physical - and, hence, nonrational - conditions of attribution. The attribution condition C_1 , as a physical attribution condition is a reflection of physical concepts. But relations between physical concepts, and, hence, between

42. "Psychophysical Laws", op. cit., p. 378.

physical attribution conditions, are not regulated by principles of rationality. Therefore, the mental property m_1 receives conditions of attribution which are not governed by the norms of rationality. Since Davidson thinks that being subject to the norms of rationality is the essence of intentional mental states, the property m_1 cannot be regarded as being an intentional mental state; it is not a mental property.

If the above is a correct interpretation of Davidson's argument, then the modality which attaches to (3) is vital if his argument is to succeed. The reason is that if there was no modality involved, and we had simply the de facto statement ' $C_1 \rightarrow m_1$ ', then C_1 could not be construed as stating a criterion of application of m_1 . Kim writes; "A de facto conditional like this cannot be taken as stating an attribution condition of m_1 no matter how loosely we construe the notion of attribution condition."⁴³ According to Davidson, attribution of mental states is essentially subject to principles of rationality. It is only if such attribution were essentially subject to some other principles would there be a possibility of a genuine clash of principles, a clash serious enough to jeopardize the ruling constitutive principle of rationality. But if the attribution of mental states, were *essentially* subject to these other principles, then there must be a *necessary* connection between those states and the attribution conditions which are partly determined by those other principles. Therefore, if the

43. "Psychophysical Laws", op. cit., p. 379.

connection between C_1 and m_1 is to be capable of undermining the constitutive principle of rationality, that connection must be necessary.

Davidson's argument (as represented by Kim) trades on modality in two places. Firstly, as the above remarks are intended to show, there must be a necessary connection between C_1 and n_1 . Secondly, the view of laws operative in the argument must be distinctly non-Humean: there must be a necessary connection between n_1 and m_1 . If we remove the modality from either (1) or (2), then we can no longer infer (3), but only its non-modal analogue ' $C_1 \rightarrow m_1$ '. But such a de facto conditional would exert no pressure on the constitutive principle of rationality, for as was argued above, a de facto conditional like this cannot be taken as stating as attribution condition of n_1 .

4

This section will attempt to identify and resolve various deficiencies in Davidson's argument. Of course, the perceived shortcomings are problems for Davidson's argument *as represented by Kim*. However, as I do not see any alternative to Kim's interpretation of Davidson, I will regard this account as correct, and henceforth speak of it as Davidson's argument.

If the schema (1)-(3) is an accurate representation of Davidson's argument, then modal connections are required in two places: In the

connection between C_1 and n_1 and in the connection between n_1 and m_1 . The question arises as to the type of modality involved in each case. In (2) $\Box[n_1 \rightarrow m_1]$ the type of modality depends on one's views of the relation between mental and physical properties: metaphysical necessity seems perhaps the most natural contender here.

However, when we begin to consider (1) $\Box[C_1 \rightarrow n_1]$ the whole notion of a condition of attribution becomes problematic. C_1 , as an attribution condition, will typically be a property or collection of properties. This property will have associated with it various concepts, and it is in virtue of these associated concepts that it can act as an attribution condition. The problem lies with the necessary connection which must exist between C_1 and n_1 if the argument is to work. For the notion of an attribution condition has epistemic connotations - it is an evidential notion. That is, one, and perhaps the most natural, interpretation of an attribution condition is a condition which states when we would be rationally justified in asserting that a given property existed. And such a condition might hold even when that property does not obtain. Given our finite capacities and limited knowledge we might have sufficient evidence for the attribution of a given property P even when such an attribution would be wrong. Our use of inductive principles of reasoning provides an example here. A closely related point is that, given our finite capacities and limited knowledge, we might wrongly, but rationally, attribute a property P on the basis of data which we wrongly, but rationally, mistook for genuine evidence. Therefore, to the extent that the notion of an attribution condition can

be given an epistemic interpretation, the connection between C_1 and n_1 is an evidential connection.

If the notion of an attribution condition is given an epistemic interpretation, the only necessary connection evident in (1) would be between C_1 and the warranted attribution of n_1 , (where warranty does not imply veracity). Would this connection be sufficient for the argument against psychophysical laws? Well, it seems the only way it could be sufficient would be if one adopted some form of instrumentalist account of physical states, according to which they are exhausted by their warranted attributions made within the bounds of some physical theory. If one wants to reject this instrumentalist position, however, one must seek an alternative interpretation of the notion of an attribution condition. For, once we allow the reality of physical states, once, that is, we allow that they transcend their warranted attributions, the connection between C_1 and n_1 , if the notion of an attribution condition is read epistemically, is a *defeasible* connection. Evidential connections between real states, states which transcend their evidential sources, are always defeasible.

To do the work required of it within the framework of a realist account of mental states, the notion of an attribution condition must be given a nonepistemic interpretation. To record this fact, I will henceforth replace talk of attribution conditions with the notion of an *individuation condition*. The remainder of this section will be concerned with characterizing this notion. With this shift, the

emphasis will swing from the conditions under which we *attribute* mental states to the conditions under which they *exist*.

An individuation condition is to be understood as an attribution condition purged of all its epistemic connotations. It is to be understood, that is, as a metaphysical counterpart of an attribution condition. An individuation condition may also provide an epistemic link with the property whose individuation condition it is, but this epistemic connection is not an essential part of what it is to be an individuation condition. There is a necessary connection between an individuation condition and its corresponding property, but the form of this necessity is *metaphysical*, and does not imply any epistemic or evidential connection.

A second important feature of individuation conditions, which they share with attribution conditions, is that they need provide only a *sufficient* condition for the existence of their corresponding property. A given property can have more than one individuation condition. In the case of attribution conditions, the dependency expressed in (1) is of only a conditional nature. So it is if we reinterpret Davidson's argument in terms of the notion of an individuation condition: the argument goes through with only a conditional, though necessary, dependence of physical property n_1 on its individuation condition C_1 . Thus, the notion of an individuation condition is to be distinguished from that of an *identity condition*, where this later notion will be understood as expressing both a sufficient *and* a necessary condition for the existence of the corresponding property.

Good examples of individuation conditions are to be found in the relation between higher level kinds and their lower level realizations. For example, the natural kind hydrogen is made up of, or realized by, three different isotopes: protium, deuterium and tritium. Each of these can be said to be an individuation condition of hydrogen. Each is sufficient, though none is individually necessary for the existence of hydrogen. So, although each individual isotope cannot be identified with hydrogen - that is, cannot be said to provide an identity condition for hydrogen - each isotope is an individuation condition of hydrogen. Moreover, the connection between each isotope and hydrogen is *modal*. And the type of modality involved here is *metaphysical*. There is no direct epistemic link between protium/deuterium/tritium and hydrogen; the relation between the three isotopes and hydrogen is discoverable only a posteriori.

Now consider the natural kind *water*. The chemical formula H_2O can be thought of as a statement of the *identity* condition of water. In virtue of hydrogen having three distinct realizations, H_2O will also. We can dub these PH_2O , DH_2O , and TH_2O . Each of these provides an individuation condition of water: a metaphysically necessary condition which is sufficient for the existence of water. Note the difficulty with the terminology here. Confusion would be generated if we write 'an individuation condition is a metaphysically *necessary sufficient* condition'. This, I think, is basically a problem of scope. We must distinguish between properties and relations holding between them. Property A provides a sufficient condition for property B, but the metaphysical necessity lies in the relation between them. What we are

trying to do in the statement of what an individuation condition is, is to attribute metaphysical necessity to the *relation* between an individuation condition and the property it individuates. This relation is one of conditional dependency, that is, a sufficient condition. To record this ambiguity of scope I will henceforth refer to an individuation condition as a metaphysically *modal* sufficient condition.

The relation between higher and lower level kinds need not be the only examples of individuation conditions. Generally: an individuation condition is to be found wherever there is a metaphysically necessary conditional dependency of one property on another. That is; property A is an individuation condition of property B if and only if B has a metaphysically modal sufficient condition in A. However, one serious qualification is called for here; the relation of determination holding between the property which provides the individuation condition and the property which is thus individuated should be *synchronic* rather than *diachronic*. For example, on an certain view of properties [i.e. the 'causal view']⁴⁴, it can be argued that the causal relations are metaphysically necessary. Thus, on this view, and without the synchronicity restriction, a cause would provide a metaphysically modal sufficient condition for its effect (more precisely, the property instantiated by the cause provides a metaphysically modal sufficient condition for the property instantiated by the effect. By 'property'

44. See, for example, Sydney Shoemaker, "Causality and Properties", in Peter Van Inwagen" ed., Time and Cause (Dordrecht, D. Reidel, 1980), pp. 109-35. Reprinted in Shoemaker, Identity, Cause, and Mind (Cambridge, Cambridge University Press, 1984), pp. 206-33.

here I mean, of course, causally powerful property). Thus, if we fail to distinguish synchronic and diachronic determination, the (causally powerful) property instantiated by a cause can be an individuation condition of the (causally powerful) property of the effect of that cause. Hence, to exclude this sort of situation, the type of determination relevant to individuation is restricted to synchronic determination.

Therefore, we arrive at the following characterization of the notion of an individuation condition:

(I) Property A is an individuation condition of property B if and only if B has a synchronic, metaphysically modal sufficient condition in A.

With this characterization of an individuation condition we can now reformulate the basic argument against strict psychophysical bridge laws:

Psychophysical bridge laws are impossible because such laws would transmit the individuation conditions of the property referred to by the reducing predicate (the physical property) to the property referred to by the reduced predicate (the mental property). That is, psychophysical bridge laws are impossible because they would give mental properties physical conditions of individuation.

This formulation preserves the spirit of Davidson's arguments and, because of a certain amount of *de-epistemologization*, avoids the major pitfalls of those arguments. The formulation will, in fact, have to be modified by way of a restriction on the types of physical properties, and consequently, a restriction on the types of individuation conditions involved. It will be seen that not just any type of physical individuation condition, transmitted to a mental state, is incompatible with the anomalism of that state. The restriction will involve introduction of the notion of *naturalness*. The reasons for this introduction will not become clear until after the following discussion of supervenience.

The Supervenience of Mental Properties: Motivation

1

In these sections, three reasons for accepting the claim that mental properties supervene on physical properties will be advanced. The first reason, to be discussed in this chapter, concerns what will be known as *general metaphysical orientation*.

The general metaphysical orientation in question leads to a view often termed *the sovereignty of physics*. I will understand this principle as making a claim primarily about properties. And with respect to properties, the sovereignty of physics makes an ontological claim best stated in terms of supervenience: once all physical properties and relations have been fixed, then all other properties and relations have been fixed. The primary aim of the sovereignty of physics thesis is to rule out the possibility of non-physical properties varying independently of physical properties. This does not deny the existence of properties which supervene on physical properties, but it does entail that there can be no dimension of freedom in what these properties do.

The specific application of the sovereignty of physics thesis to mental properties is well brought out by a passage from McGinn.

" In the case of many mental states (though perhaps not all) it appears evident that a difference of mental state (uncompensated by other mental differences) implies a difference of behavioural dispositions; and it also seems undeniable that a difference between the dispositions of two animal bodies depends, other things being equal, upon a difference in their brain states. Suppose then that two creatures had totally different mental attributes, though their brain states were exactly the same. That would imply that their dispositions to behaviour were correspondingly different, even though their brain states differed not at all. But this would be to deny that the brain states of human and animal bodies were responsible for how those bodies are apt to behave. Therefore we must reject the original supposition that mental properties are independent of physical properties. The thesis that the mental is so determined by the physical is sometimes called the supervenience thesis: it holds that mental properties cannot vary while physical properties are kept constant."⁴⁵

2

The sovereignty of physics is, and always will be, the principle reason for believing that mental properties supervene on physical properties. However, the following sections will discuss more specific reasons for adopting the supervenience claim. The discussion will focus on the connection between supervenience and anomalism, and, accordingly, will centre around the question of what consequences the principle of the anomalism of the mental has for the claim that mental properties supervene on physical properties.

45. Colin McGinn, The Character of Mind (Oxford, Oxford University Press, 1982), p. 29.

Anomalous monism has been charged with leading to epiphenomenalism.⁴⁶ Consider, for example, the following passage from Kim:

" Consider Davidson's account: whether or not a given event has a mental description (optional reading; whether it has a mental characteristic) seems entirely irrelevant to what causal relations it enters into. Its causal powers are wholly determined by the physical description or characteristic that holds for it; for it is under its physical description that it may be subsumed under a causal law. And Davidson explicitly denies any possibility of a nomological connection between an event's mental description and its physical description that could bring the mental into the causal picture."⁴⁷

46. For example, see: T. Honderich: "The Argument for Anomalous Monism", Analysis, vol.42, no.1, (1982), pp. 59-64. "Psychophysical Lawlike Connections and Their Problem", Inquiry, vol.24, no.3, (1981) pp. 277-304. "Causes and if p, even if x, still q", Philosophy, vol. 57, (1982), pp. 291-317. "Smith and the Champion of Mauve", Analysis, vol.44, no.7, (1982), pp. 86-89. Peter Smith's replies to Honderich are contained in: "Bad News for Anomalous Monism?", Analysis, vol.42, no.7, (1982), pp. 220-24. And "Anomalous Monism and Epiphenomenalism: A Reply to Honderich", Analysis, vol.44, no.2, (1984), pp. 83-86. Other sources of the charge of epiphenomenalism are the following: F. Stoutland: "The Causation of Behaviour", in Essays in Honour of G.H. Von Wright (Acton, Philosophica Fennica, 1976), pp. 286-325. And "Davidson on Intentional Behaviour", in LePore and McLaughlin eds., Actions and Events: Perspectives on the Philosophy of Donald Davidson op. cit., (1985), pp. 44-59. Jaegwon Kim: "Causality, Identity, and Supervenience in the Mind-Body Problem", Midwest Studies in Philosophy, vol.4, (1979), pp. 31-49. "Supervenience and Supervenient Causation", The Southern Journal of Philosophy vol. XXII, ed., T. Horgan, (The Spindel Conference 1983: Supervenience), pp. 45-56. "Epiphenomenal and Supervenient Causation", Midwest Studies in Philosophy vol.9, (1984), pp. 257-70. H. Robinson: Matter and Sense, Cambridge, Cambridge University Press, 1982), ch.1, pp. 8-19. E. Sosa: "Mind-Body Interaction and Supervenient Causation", Midwest Studies in Philosophy vol.9 (1984), pp. 271-81. D. Follesdal: "Causation and Explanation: A Problem in Davidson's view on Action and Mind", in LePore and McLaughlin eds., Actions and Events: Perspectives on the Philosophy of Donald Davidson, op. cit., pp. 311-23. M. Johnston: "Why Having a Mind Matters", in LePore and McLaughlin eds., Actions and Events: Perspectives on the Philosophy of Donald Davidson, op. cit., pp. 408-26.

47. Kim: "Epiphenomenal and Supervenient Causation".

The above claim is ambiguous; two positions need to be distinguished: (a) epiphenomenalism with respect to *events*, which corresponds to the claim that mental events are causally inefficacious, and (b) epiphenomenalism with respect to *properties*, which corresponds to the claim that mental properties have no causal powers. These theses are plainly distinct: one can hold that mental events are causally efficacious without holding that mental properties have causal powers. The principles involved here can best be discussed by reference to the more explicit argument given by Honderich.⁴⁸

(a) In any causal transaction involving events, there are some properties of those events which are relevant and some which are irrelevant to that transaction. An event is the cause or effect it is in virtue of only certain of its properties.

(b) If A caused B in virtue of a property f of A and a property g of B, then A and B are in lawlike connection in virtue of the properties f and g. That is, causally relevant properties are nomic - 'The Principle of the Nomological Character of Causally Relevant Properties'.

(c) Now, if mental events cause physical events, they must do so in virtue of certain of their properties.

(i) The claim that these properties are *mental* is incompatible with the denial of psychophysical laws, for, by (b), all causally relevant properties are nomic.

48. "The Argument for Anomalous Monism", op. cit.

(ii) The claim that the relevant properties are *physical* undermines the efficacy of the mental, for the "mentality" of any given mental event would appear to cease being relevant to its causal interactions with other events.

The first point to note is that Honderich makes no principled distinction between causal *relevance* and causal *power*. Consequently, his use of the expression 'causal relevance' diverges from my use of that term. If one adheres to the usage defined in this thesis, then whether one understands epiphenomenalism as a claim about causal power or as a claim about causal relevance depends upon whether one understands it as making an ontological claim about dispositions to causal efficacy, or as an epistemological claim about which properties play an explanatory role in (causal) explanations. That is, on the ontological reading, the claim that mental properties are epiphenomenal is the claim that they do not bestow on their instances dispositions to enter into token causal relations. On the epistemological reading the claim is that mental properties play no explanatory role in explanations. Honderich does appear to understand epiphenomenalism as primarily an ontological thesis concerning causal powers. Accordingly, I will reinterpret his references to causally relevant properties in terms of causally powerful properties. However, for the purposes of this chapter, the distinction between causal power and causal relevance is of only secondary importance. Consequently, if desired, reinterpretation of the arguments in terms of causal relevance can be made without much revision.

Of far more importance in the assessment of these arguments is the distinction between types and tokens and, consequently, the distinction between causal power and causal efficacy. The first problem with Honderich's argument is deciding whether it is meant to establish event or property epiphenomenalism. Indeed, Honderich seems to conflate the two theses: the claim that mental events are causally inefficacious, and the claim that mental properties have no causal powers. Honderich conflates these two claims because he conflates types and tokens, properties with their instances. Therefore, if the argument is to be properly evaluated, the distinction between properties and their instances must be re-established.

Insofar as Honderich is arguing for event epiphenomenalism, that is, insofar as he is arguing for the causal inefficacy of mental events, his argument can be refuted by the following line of reasoning. In what follows, I will, for expository purposes, freely talk of events as instances of properties, even though this only applies to a small subclass of events, other events being composed of instances of properties. Nothing essential to the argument will be lost by this simplification.

If Honderich can be correctly interpreted as attacking the possibility of combining the anomalism of the mental with a view of mental events as causally efficacious, then he seems to forget that it is properties instantiated which are causally efficacious; it is instances of properties, not the properties themselves, which have causal efficacy. Properties can have causal power, they can have causal relevance; but they cannot have causal efficacy. It is dated,

particular tokens (instances of properties) which have causal efficacy. This can be seen from such well known facts as that whilst event-tokens enter into causal relations under descriptions (i.e. picked out by properties), the same event-token can enter into the same causal relation irrespective of the description, or the properties, by which it is picked out.

Under the modification explained earlier, Honderich's 'Principle of the Nomological Character of Causally Relevant Properties' alters to the 'Principle of the Nomological Character of Causally Powerful Properties'. And this principle cannot be understood as a principle pertaining to causal *efficacy*. It cannot be understood, that is, as a principle governing causal relations, where these are understood as dated, particular relations holding between dated, particular individuals. Rather, it must be a principle governing what sorts of *properties* can have causal *power*. Its claim is that all such properties are nomic.

If Honderich is making a point about mental *event* epiphenomenalism, then he needs a principle which governs relations between events, or instances of properties. And this principle would have to be stated thus: 'The principle of the Nomological Character of Causally Efficacious *Instances* of Properties.'⁴⁹ This states that if two instances of properties are related as cause and effect, they must be

49. c.f. C. Macdonald & G. Macdonald: "Mental Causes and the Explanation of Action", in The Philosophical Quarterly, vol.36, no.143, pp. 145-158, p. 148.

instances of some nomic properties. It is this principle which Honderich requires if he wishes to make a point about event causation. Hence, it is this principle he requires if he wishes to argue for mental event epiphenomenalism. But if he accepts this principle, his argument fails.

The reason is that any instance of a property may be an instance of other properties also. Ali's punching of Frazier may be an instance of the property, being a punching, but it may well also be an instance of the property, being a movement of the arm, or being a knocking out of the champion, or being a winning of the world heavyweight boxing title. By 'The Principle of the Nomological Character of Causally Efficacious Instances of Properties', events bearing causal relations to one another must be instances of properties referred to by expressions occurring in causal laws. But this does not mean that every property instanced by causally related events must occur, or be referred to, in a law.

Ali punches Frazier. An event which is an instance of the property being a desire to punch Frazier causes an event which is a movement of the arm. But since an event can be an instance of two or more distinct properties, it may well be the case that one and the same event is both an instance of the property, being a desire to punch Frazier, and an instance of another, physical, property, say being a brain event *b*, where being an instance of the former just is being an instance of the latter. The two properties share the same instance. Thus, the instance of the property, being a desire to punch Frazier (a mental property) will satisfy the principle of the nomological character of causally

efficacious instances of properties because this instance is *identical* with an instance of a physical property which is nomic.

Therefore, Honderich's arguments cannot be taken as establishing mental event epiphenomenalism. However, perhaps Honderich is to be interpreted as arguing for mental property epiphenomenalism. That is, perhaps he is to be interpreted as arguing for the claim that mental properties have no causal powers. In this case, the principle required is indeed the principle of the nomological character of causally powerful properties. This states that every causally powerful property is nomic.

The problem with this principle, however, is that it does not seem very plausible. That is, it does not appear that all properties of causal power are nomological. A knife, for example, has various causal powers - being able to cut through butter etc., etc. And this causal power cannot plausibly be regarded as nomological. The causal powers of such everyday objects as knives, rocks, chairs, feathers, etc, do not, intuitively, seem to be nomological. So either these types of entity do not have causal power, or Honderich's principle must be rejected. And it is clear that the latter course should be adopted.

In the case of most macro-physical objects and properties, we can say that their possession of causal powers does not imply the nomologicality of those powers. However, the possession of these powers is dependent upon possession of other properties which are nomological in the true sense of that word. The macro-objects possessing macro-

properties will be composed of micro-objects possessing micro-properties. And the causal powers possessed by these micro-properties will be nomological. The causal powers of the macro-properties will then be supervenient on the nomological causal powers of the micro-properties. And such supervenience is all that seems to be required for possession of causal power. The supervenience will typically be implemented by the possessors of the non-nomological causal powers themselves supervening on the possessors of the nomological causal powers. On this view, non-nomological causal powers are essentially *dependent* causal powers, but they are genuine causal powers nonetheless.

3

The arguments of the previous section strongly suggest that if one accepts the principle of the anomalism of the mental, then one needs a supervenience thesis to secure the causal powers of mental properties. And, this claim will henceforth be regarded as correct. However, it will now be argued that matters might be yet more complicated. In particular, it will be argued that the supervenience of mental on physical properties *might* follow directly out of the claim that mental events are identical with physical events. At least, it follows from one view of event identity. If this is so, then given the correctness of the above remarks, mental-physical property supervenience is a necessary condition of avoiding *event* epiphenomenalism (given anomalism).

The conception of an event as a non-decomposable basic particular capable of satisfying various non-equivalent descriptions plays an important role in Davidson's articulation of a token identity theory of mind and body. According to Davidson, the identity of a mental event with a physical event will be a matter of the event satisfying both mental and physical descriptions.⁵⁰ However, if we allow that events typically enter into relations of composition with one another,⁵¹ and that the events typically involved in issues of mind body identity are composite, Davidson's theory houses an important ambiguity. This is, perhaps, best revealed by transposition (insofar as this is possible) into the property-exemplification idiom.

The property-exemplification account, baldly stated, views events as exemplifications of properties (event-types) by objects at times. That is, events are *property-instances*.⁵² If we accept the composition principle, then an event, typically, will be a property-instance which is composed of other property-instances. Now, given this conception, Davidson's account of token identity does not possess the resources sufficient to distinguish the following two cases.

Case (1) consists of an event E which is composed of several property-instances. One of these instances, m_1 , is an instance of a

50. "Mental Events", op. cit. p. 224.

51. See Lombard, "Events", op. cit., and Events: A Metaphysical Study, op. cit.

52. See Kim, "On the Psychophysical Identity Theory", op. cit., "Causation, Nomic Subsumption and the Concept an Event", op. cit., and "Events as Property Exemplifications", op. cit. Also, Lombard, "Events", op. cit., and Events: A Metaphysical Study, op. cit.

mental property M, and another instance, p_{E1} , is an instance of a physical property P. In this case, m_1 and p_1 are *distinct* instances which *co-occur* in the same event E. However, their distinctness does not preclude Davidson's criterion of token identity being satisfied. Mental and physical descriptions based on M and P apply to the same event E. E satisfies both a mental and a physical description in virtue of M and P being *co-instantiated*. And M and P are co-instantiated in virtue of their instances, m_1 and p_1 , co-occurring.

However, we can also consider a quite distinct case, (2), which also satisfies Davidson's condition on mental/physical token-identity. Case (2) consists in an event E' which is also composed of several property-instances. One of these, m_1' is an instance of a mental property M'. In this case m_1' is *identical* with an instance p_1' of physical property P'. Mental and physical descriptions, here, do apply to E'; for properties M' and P' are co-instantiated in E'. However, the relation between M' and P' is much stronger than mere co-instantiation. For M' and P' share the same instance: m_1' and p_1' do not merely co-occur, they are identical. M' and P' share the same instance, and this is a much stronger relation than that of being merely instantiated in the same thing.

Davidson's conception of token identity is based around the idea of mental and physical descriptions applying to the same thing. And this idea can be cashed out in terms of co-instantiation of properties. The problem is that Davidson's account provides no resources for

distinguishing co-instantiation due to *co-occurrence* of instances, and co-instantiation due to *identity* of instances.

Whether one adopts (1) or (2) as the proper characterization of mental/physical token identity is not an idle question. The decision will be important with regard to the relation thought to hold between mental and physical properties.

Consider, first, the identity theory characterized by (1). An analogy would be the case of two properties - say, the property of being red and the property of being square - being instantiated in the same object; a red square. In this case, the instance of redness and the instance of squareness co-occur in the same object. And it is fairly clear that such co-occurrence can obtain without there being any determinate relation between the properties whose instances co-occur. Thus, the property of being red and the property of being square are mutually independent, and this does not preclude their instances co-occurring in the same object. By analogy, then, we should expect that a mind-body token identity theory based on co-occurrence of instances would, by itself, have no consequences for any relation holding between mental and physical properties.⁵³

In many ways, the model of identity based around co-occurrence of property-instances does not seem to be an adequate characterization of

53. This view of mind-body token identity is clearly implicated in, for example, McGinn, The Character of Mind, op. cit., p. 29.

mind-body token identity. In particular, it seems possible for the condition of co-occurrence of property-instances to be satisfied within some broadly dualist framework. Even Cartesian dualism maintains co-occurrence of mental and physical events occurring inside the head. It seems possible for the Cartesian to allow that at any time inside the head there will be a mental event co-occurring with a physical event where such co-occurrence takes place in a gerrymandered conjunction of mental and physical substances.

Whether or not this latter contention is ultimately correct, the situation is very different for the token identity theory which asserts the identity of mental and physical property-instances. It seems that any mutual independence of mental and physical properties would prevent their sharing an instance. Consider, first, the case of mutually *contrary* properties; identity of instances, here, seems impossible. An instance of redness cannot be identical with an instance of blueness. And in the case of mutually independent properties, the point seems almost as clear. Intuitively, it does not seem possible for an instance of redness to be identical with an instance of squareness. While it is not difficult to see how instances of these properties might co-occur, it is difficult to see how they might be identical.⁵⁴

54. It is almost certainly considerations such as these which motivate Kim's criterion of event identity. According to this, $\text{event}(x,P,t) = \text{event}(y,Q,t')$ if and only if $x=y$, $P=Q$, and $t=t'$. The considerations described here motivate the condition of property identity. It will be argued that this requirement is too strong, and rests on a type-token confusion.

It seems that if two properties are to share a single instance, there must be some fairly specific relation between them. According to Kim, the relation must be that of *identity* of properties. Thus, he proposes the following as an identity condition for events:

Event [x,P,t] = Event [y,Q,t'] just in case x=y, P=Q, t=t'.

where x, P, and t are constitutive objects, properties, and times respectively of the event. Thus, according to Kim, property identity is a precondition of identity of instances of those properties; token identity implies type identity.⁵⁵

I think this condition is too strong; in particular, identity of properties is not a necessary condition of identity of their instances. This may appear counterintuitive. How can x's having P at t be identical with x's having Q at t if P≠Q? However, I believe the apparent implausibility here results from failure to adequately distinguish properties and their instances. Once again, the comparison with objects proves illuminating.

Consider the relation between the property of being red, and the property of being coloured. Possession of the former entails possession of the latter but not vice versa, so the properties are plainly distinct. Even so, no one would want to claim that an object's

55. "On the Psychophysical Identity Theory", op. cit., "Causation, Nomic Subsumption, and the Concept of an Event", op. cit., and "Events as Property Exemplifications", op. cit.

exemplification of a colour, say redness, requires first that it contain an instance of the property of being red, and then that it contain a distinct instance of the property of being coloured. The instance of the former, in this case, just is (i.e. identical with) the instance of the latter.⁵⁶

Here, the relation between the properties which makes possible their sharing of a common instance seems to be the relation of *entailment*. However, there seems no reason to restrict this notion to its purely logical sense; the notion of necessary determination (with the force of the modality perhaps varying from case to case) should do equally as well. If this is correct, Kim's identity condition for events needs to be weakened to:

Event [x,P,t] = Event [y,Q,t'] just in case, $x=y$, $t=t'$,
and $P=Q$, or $P \rightarrow Q$.

where ' \rightarrow ' is to be understood as entailment in this broader sense. Note also that the notion of determination at work must be of a synchronic, rather than a diachronic, character. This is necessary to safeguard the condition $t=t'$.

In the mind-body case, the relation of necessary synchronic determination is provided by the claim that mental properties supervene

56. c.f. C. Macdonald & G. Macdonald, "Mental Causes and the Explanation of Action", op. cit., p. 149.

on physical properties. If this is right, then mind-body token identity, understood as asserting identity of property-instances rather than mere co-occurrence, entails that mental properties supervene on physical properties. Token identity, in this sense, implies property supervenience.

4

If one accepts the principle of the anomalism of the mental, then, a supervenience thesis is needed in order to maintain that mental properties have causal powers. That is, it is needed in order to avoid the charge of mental property epiphenomenalism. Moreover, given a view of event identity derived from a property-exemplification account of events, a commitment to supervenience is essential to avoid the charge of mental event epiphenomenalism. For, the reply to this charge turns upon the assumption that mental events can have causal efficacy because they are identical with physical events. And if the remarks of the previous section are correct, on the view of event identity inspired by a property-exemplification account, mental-physical event identity entails mental-physical property supervenience.

The Supervenience of Mental Properties: Characterization.

1.

In the following passage, Davidson introduces the concept of supervenience as an account of the relation between mental and physical properties:

" Although the position I describe denies there are psychophysical laws, it is consistent with the view that mental characteristics are in some sense dependent, or supervenient, on physical characteristics. Such supervenience might be taken to mean that there cannot be two events alike in all physical respects but differing in some mental respect, or that an object cannot alter in some mental respect without altering in some physical respect. Dependence or supervenience of this kind does not entail reducibility through law or definition."⁵⁷

Davidson here gives two formulations of the supervenience principle: As the claim that, (i) "... there cannot be two events alike in all physical respects but differing in some mental respect," and as the claim that, (ii) "... an object cannot alter in some mental respect without altering in some physical respect." These formulations are non-equivalent. It will be seen that most formulations of the supervenience

57. "Mental Events", op. cit., p. 214. Davidson's most detailed discussion of the supervenience relation is to be found in "The Material Mind", op. cit.

principle present in contemporary literature conform to one or other of Davidson's statements.

These sections will consist in an examination of the concept of supervenience as it appears in recent literature. This examination will be developed along two axes corresponding to, on the one hand, the logical form of the supervenience principles, and, on the other hand, the strength or range of application of such principles. Accordingly, the first stage of the examination will be concerned with delineating the logical forms of the various formulations of supervenience. The second stage will be concerned with the modal status of the supervenience principle: it will be concerned with the propriety and placement of modal operators with respect to the statements of logical form.

2

Jaegwon Kim distinguishes two forms of supervenience, "Weak" and "Strong".⁵⁸ This distinction turns purely on the modal status of the two principles, that is, it turns purely on the placement of modal operators. And abstracting from this placement, the logical form of the two principles is identical. This section will focus purely on this

58. "Concepts of Supervenience", Philosophy and Phenomenological Research, vol. XLV, no.2, pp. 153-76. Also "Supervenience and Supervenient Causation", The Southern Journal of Philosophy, vol. XXII Supplement, ed., T. Horgan, (The Spindel Conference 1983: Supervenience), pp. 45-56.

logical form. The discussion of the modal status of the two principles, and, hence, of the distinction between the two principles, will be postponed to a later section.

Let A and B be two non-empty families of properties closed under the usual Boolean property-forming operations - conjunction, complementation and disjunction. Kim defines *weak supervenience* as follows:

" A *weakly supervenes* on B just in case necessarily for any x and y, if x and y share all properties in B then x and y share all properties in A - that is, indiscernibility with respect to B entails indiscernibility with respect to A"⁵⁹

A is the supervenient family, B is the supervenience base. Suppose the set A consists of the single property of being a good man (G). Closed under Boolean operations it will contain two properties; G and -G (besides the tautological property $G \vee \neg G$, and the impossible one $G \& \neg G$). Set B contains the properties of being courageous (C), being benevolent (V), and being honest (H). If A weakly supervenes on B, then if two men share the same properties in B, say both are honest and benevolent but lack courage, then both must be good men, or neither is. Any difference in A must be accounted for by some difference in B.⁶⁰

59. "Concepts of Supervenience", op. cit., p. 158.

60. "Concepts of Supervenience", op. cit., p. 158-59.

The closure of B under Boolean construction operations will (again excluding tautological and impossible properties) consist of what Kim calls "B-maximal properties". These are the strongest consistent properties constructible in B. In this example there are eight: C&V&H, C&V&-H, C&-V&H, -C&-V&-H. These properties are mutually exclusive, and every object must have just one of them. Two objects are indiscernible in B just in case they have the same B-maximal property. Weak supervenience of A on B, therefore, is the thesis that any two objects with the same B-maximal property must have the same properties in A (i.e. G or -G). There is no possible world in which two objects have the same B-maximal property and yet differ in respect of G.

The logical form of Kim's version of weak supervenience is this:

$$\Box (x)(y) [(G)(Gx \leftrightarrow Gy) \rightarrow (F)(Fx \leftrightarrow Fy)] \quad (1)$$

where x, y are objects.

G is a B-maximal property.

F is a supervenient property (i.e. a property of the supervenient family).

The discussion of the attached modal operator will be postponed to a later section. What is of interest here is the form of the principle irrespective of its modal status.

Terence Horgan presents two formulations of the supervenience principle:⁶¹

" There do not exist any two world-regions which are exactly alike in all qualitative features of kind J, which differ in some qualitative intrinsic feature of kind K, and which belong respectively to two possible worlds each satisfying accessibility condition C."

" There do not exist any two world-regions, belonging respectively to two possible worlds each satisfying accessibility condition C, containing two individuals which are exactly alike in all qualitative intrinsic properties of kind J but different in some qualitative intrinsic property of kind K."⁶²

Abstracting from the notion of an accessibility condition, it is possible to arrive at the following common logical form:

$$- (Ex)(Ey)[(J)(Jx \leftrightarrow Jy) \rightarrow (EK)((Kx \& - Ky) \vee (Ky \& - Kx))] \quad (2)$$

And this is equivalent to Kim's formulation.

John Haugeland presents what he calls "weak supervenience".⁶³ Where W is a set of possible worlds he stipulates that "... two worlds in W are discernible with language L just in case there is a sentence of L which is true at one, and not true at the other." He then defines weak supervenience as a relation between two languages K and L relative to W.

61. "Supervenience and Microphysics", Pacific Philosophical Quarterly, vol.63, pp. 29-43.

62. "Supervenience and Microphysics", op. cit., p. 40.

63. "Weak Supervenience", American Philosophical Quarterly, vol.19, pp. 93-103.

" K weakly supervenes on L (relative to W) just in case any two worlds in W discernible with K are discernible with L."⁶⁴

The logical form of Haugeland's formulation is this:

$$(K)_{\omega}(L)_{\omega}(x)(y)[Kx \ \& \ -Ky \ \rightarrow \ Lx \ \& \ -Ly] \quad (3)$$

The formulations of Kim and Horgan follow Davidson's first formulation of the supervenience principle: there cannot be two events alike in all physical respects but differing in some mental respect. That is, physical indiscernibility entails mental indiscernibility. Haugeland's account, however, follows Davidson's second formulation, an object cannot change in some mental respect without also changing in some physical respect: no difference without a physical difference.

Hellman and Thompson also formulate a version of supervenience which follows Davidson's second formulation: an object cannot alter in some mental respect without altering in some physical respect. The Hellman-Thompson account is formulated in terms of predicates and can be stated thus⁶⁵:

$$(S)(u)(v)(EP)(Su \ \& \ -Sv \ \rightarrow \ Pu \ \& \ -Pv) \quad (4)$$

64. "Weak Supervenience", op. cit., p. 97.

65. G. Hellman and F. Thompson, "Physicalism: Ontology, Determination, and Reduction", The Journal of Philosophy, vol.72, pp. 551-64.

where u, v are n -tuples of objects. P ranges over physical predicates (subvenient predicates). S ranges over non-physical predicates (supervenient predicates).

This formulation, and that of Haugeland, are stated in terms of predicates applying to different objects at the same time (or, in Haugeland's case, of languages applying to worlds at given times). However, their connection with the second of Davidson's formulations can be exhibited by switching from talk of predicates to talk of properties, and instead of considering two different objects, considering one object at different times. Then, the following formulation can be arrived at.

$$(F)(x)(t)(EG)(Fx_t \ \& \ - \ Fx_{t^*} \ \rightarrow \ Gx_t \ \& \ - \ Gx_{t^*})$$

where F ranges over supervenient properties,

G ranges over subvenient properties,

x ranges over objects,

t ranges over times. $t \neq t^*$.

Therefore, there are two basic formulations of the principle of supervenience present in recent literature. On the one hand, following Davidson's first formulation, there are the accounts of Kim and Horgan. On the other, following Davidson's second formulation are the accounts of Haugeland and Hellman and Thompson. The two types of formulation are not equivalent. However, both seem to be accurate and reasonable statements of the supervenience principle. Nothing much turns on which formulation is taken to be more basic, or more correct. For reasons

largely to do with personal preference, I will focus on the account of Kim. Therefore, the principle of supervenience will, from now on, be regarded as having a logical form expressible as follows.

$$(x)(y) [(G)(Gx \leftrightarrow Gy) \rightarrow (F)(Fx \leftrightarrow Fy)]$$

3

This section will be concerned with the modal status of the supervenience principle, understood as above.

A desideratum of any characterization of the supervenience relation is that it accounts for the intuition that if F supervenes on G, then any object which becomes G-like *must* become F-like. That is, the supervenience of F on G must imply more than a de facto connection between Gs and Fs. The first consequence of this desideratum is that the supervenience principle must have application beyond merely the actual world. This explains the presence of the modal operator in Kim's formulation of weak supervenience.⁶⁶

$$\Box (x)(y) [(G)(Gx \leftrightarrow Gy) \rightarrow (F)(Fx \leftrightarrow Fy)]$$

66. See "Concepts of Supervenience", op. cit.

(where F is the supervenient property or family of properties, G is the base property or family).

Thus, Kim's statement, differs from that of Horgan (and from those of Haugeland, and Hellman-Thompson) by requiring that supervenience should hold not just in the actual world but in all possible worlds in some range of possible worlds left open to be specified as a specific application of the concept. Specifying this range involves, primarily, specifying the type of necessity denoted by the operator - logical, metaphysical, nomological etc. It is important to realize, however, that the placement of the modal operator does not call for comparisons across possible worlds, but only within any possible world. And it is this fact which indicates why Kim calls this type of supervenience "weak".⁶⁷

The weakness of weak supervenience lies in this: as defined, it prohibits the possibility of two things agreeing in base properties but differing in supervenient properties only *within* any possible world. Transworld B-indiscernibility permits transworld A-discernibility. This leaves us with a puzzle, developed by Blackburn as follows:⁶⁸ In any possible world, once there is an object which is F, and whose F-ness supervenes on G, then in that world, anything else which is G is F also. Call these G/F worlds. However, weak supervenience allows that there

67. See, "Concepts of Supervenience", op. cit.

68. Simon Blackburn, "Supervenience Revisited", in I. Hacking ed., Exercises in Analysis (Cambridge, Cambridge University Press), especially pp. 58-61. See also Blackburn's discussion of supervenience in Spreading the Word (Oxford, Oxford University Press, 1984), ch.6.

are other worlds in which things are G but not F. Call these G/0 worlds. The problem is, why there are no mixed worlds, where some things are G and F, and some things are G but not F; that is, why are there no worlds of the form G/Fv0? These worlds are ruled out by weak supervenience, but why? It is difficult to see why. After all, weak supervenience allows both G/F and G/0 situations. The problem, at root, seems to lie in the failure of weak supervenience to specify an adequate relation of determination. A desideratum for such a relation would be that it accord with the intuition that, if F supervenes on G, then if any object were to become G-like, it would become F-like. For if the relation were not this strong, it is difficult to see how much substantive content supervenience, as a relation of determination, possesses.

This point needs emphasizing. We are to suppose that G determines F in one world, a G/F world, but not in another world, a G/0 world. But, then, it becomes unclear in what sense G determines F in the G/F world. For the existence of the G/0 world seems to indicate that it cannot be G *simpliciter* determining F, but, rather, that it is something about the world in which G occurs which is also a partial determinant of F. That is, the existence of the G/0 world shows that G, by itself, cannot determine F. And then it becomes extremely unclear just in what sense the relation expressed by weak supervenience is a relation of determination at all.

The need for a second modal operator in order to give a viable account of supervenience was first pointed out by McGinn.

" His [Davidson's] formulation seems best captured by means of a condition containing two occurrences of a necessity operator, thus: necessarily, for any mental property instantiated by a creature at a time, there is a physical property the creature instantiates such that, necessarily, if any creature instantiates that physical property, then it instantiates that mental property."⁶⁹

The idea is developed by Kim as follows. Consider a definition of weak supervenience which Kim argues is equivalent to the first definition:

" A weakly supervenes on B if and only if necessarily for any property F in A, if an object x has F, then there exists a property G in B such that x has G, and if any y has G, it has F."⁷⁰

The logical form of the new definition is;

$$\Box (F)(x) [Fx \rightarrow (EG)(Gx \ \& \ (y)(Gy \rightarrow Fy)] \quad (5)$$

The key aspect of the second definition is its last clause to the effect that any object having G also has F. The force of this claim is simply that within each possible world this general G-F connection must hold. There is no requirement that this connection have transworld stability. In order to secure such stability, Kim strengthens the last clause by modalizing thus:

69. Colin McGinn: "Philosophical Materialism", *Synthese*, vol.44, pp. 173-206, p. 197.

70. "Concepts of Supervenience", op. cit., p. 163.

" A *strongly supervenes* on B just in case, necessarily, for each x and each property F in A, if x has F, then there is a property G in B such that x has G, and necessarily if any y has G it has F."⁷¹

That is;

$$\Box (F)(x) [Fx \rightarrow (EG)(Gx \ \& \ \Box(y)(Gy \rightarrow Fy))] \quad (6)$$

It is the second modal operator which secures transworld stability. The corresponding strong version of (1) will then be:

$$\Box (x)(y) [(G)(Gx \leftrightarrow Gy) \rightarrow \Box(F)(Fx \leftrightarrow Fy)] \quad (7)$$

That is; if x and y are equivalent in point of G, then necessarily (i.e. in all possible worlds) they are equivalent in point of F. (6) and (7) form two versions of Kim's strong supervenience.

Therefore, the notion of supervenience needed in order to characterize the relation between mental and physical properties will, henceforth, be understood as that expressed in (7) above. Following Kim, I will call this *strong supervenience*.

71. "Concepts of Supervenience", op. cit., p. 165.

Anomalism and Supervenience: Compatibility

1

Having decided to accept that an adequate expression of the principle of supervenience is provided by Kim's formulation of *strong* supervenience, and having accepted a characterization of this strong supervenience as;

$$\Box (x)(y)[(G)(Gx \leftrightarrow Gy) \rightarrow \Box (F)(Fx \leftrightarrow Fy)]$$

I now wish to discuss whether supervenience, as characterized, entails the reduction of supervenient to base properties. More precisely, I want to consider whether strong supervenience of mental on physical properties is incompatible with the anomalism of mental properties.

Kim is one who believes that strong supervenience is incompatible with anomalism:

" If A strongly supervenes on B then for each property F in A there exists a property G in B such that necessarily $(x)[G(x) \leftrightarrow F(x)]$. That is, every A-property has a necessary coextension in B."⁷²

72. "Concepts of Supervenience", op. cit., p. 170. c.f. "Supervenience and Supervenient Causation", op. cit., pp. 47-48.

Therefore, Kim claims, strong supervenience entails reduction via biconditional bridge laws. Kim's basic line of reasoning runs as follows: Given the strong supervenience of mental on physical properties, each mental property has a sufficient condition in some physical property which realizes it. Take any mental property m_1 . Then given strong supervenience, m_1 has a sufficient condition in some physical property p_1 which realizes it. Now, given the possibility of variable realization, m_1 can have different realizing bases; call these $p_2, p_3, p_4, \dots, p_n$. Nevertheless, it is possible to take all the realizing properties m_1 does or might have, and close them into a single property by Boolean disjunction. It does not matter that the number of potential realizing properties is infinite; for according to Boolean theory, properties can be closed under infinite disjunction. Since the connection between each realizing base and the supervenient property is necessary, it follows that the connection between the supervenient mental property and the supervenience base taken as a whole - i.e. as the disjunction of realizing properties - is also necessary. Therefore, for each mental property which supervenes on a physical property, there is a modal biconditional of the form $[m_1 \leftrightarrow (p_1 \vee p_2 \vee \dots \vee p_n)]$ relating it to a physical base.

More fully, his argument runs as follows.⁷³ Let F be a property in A . We may assume F to be contingent; that is, some x has F in some possible world w . By the definition of strong supervenience there is a

73. "Concepts of Supervenience", *op. cit.*, p. 170.

property G in B such that x has G (in w) and necessarily $\langle y \rangle [G(y) \rightarrow F(y)]$. Let $B_{x,w}$ be the B-maximal property x has in w. We have then;

$$\Box \langle y \rangle [B_{x,w}(y) \rightarrow G(y)]$$

whence,

$$\Box \langle y \rangle [B_{x,w}(y) \rightarrow F(y)]$$

And for each v that has F in a world u, we will have;

$$\Box \langle y \rangle [B_{v,u}(y) \rightarrow F(y)]$$

Let B^* be the infinite disjunction of these B-maximal properties. Then;

$$\Box \langle y \rangle [B^*(y) \rightarrow F(y)]$$

We also have the converse;

$$\Box \langle y \rangle [F(y) \rightarrow B^*(y)]$$

For suppose not. Then in some possible world $w\theta$, there is an object such that $F(x)$, but not $B^*(x)$. But by strong supervenience there is some property K in B such that $K(x)$ in $w\theta$ and necessarily $\langle y \rangle [K(y) \rightarrow F(y)]$. Let $B\theta$ be the B-maximal property of x in $w\theta$. Then, as before, $\langle y \rangle [B\theta(y) \rightarrow F(y)]$, and it follows that $B\theta$ is one of the disjuncts in

B*. Hence, x must have B*, yielding a contradiction. We therefore have;

$$\square (y)[B^*(y) \leftrightarrow F(y)]$$

Crucial to this argument is the nature of B*: B* is not a simple property, it is *doubly* composite. Firstly, it is a disjunctive property, a disjunction of simpler properties. Secondly, the disjuncts are themselves composite; they are B-maximal properties - properties constructed from other properties under Boolean conjunction and complementation.

2

The argument appears to entail that strong supervenience is just as incompatible with anomalism as psychophysical laws would be. For a modal biconditional of the form described above seems appropriate for substitution in the argument scheme given against psychophysical laws in part one of this chapter. However, appearances can be deceiving, and I shall argue that a modal biconditional *of the form described above* is not suitable for substitution in such a schema, and hence is compatible with mental anomalism.

The reason for this compatibility lies in the *composite* nature of the envisaged supervenience base properties. The property which forms the supervenience base in the above biconditional is, in fact, *doubly*

composite. For the sake of clarity, let us distinguish *supervenience base properties* from *realizing base properties*.

- (1) The *supervenience base property* of any given mental state is composed of an infinite disjunction of more simple properties. I will call these the *realizing base properties* of the supervenient mental state. The supervenience base property is thus a composite property, formed by Boolean disjunction, and purports to provide both a necessary and sufficient condition for the existence of the given supervenient mental state.
- (2) Each realizing base property purports to provide a sufficient condition for the existence of the given supervenient mental state. On Kim's account of supervenience, realizing base properties are known as 'maximal properties' and are composite, typically constructed by way of Boolean conjunction and complementation. It will be argued that this compositeness holds irrespective of whether one adopts Kim's account.

The distinction between supervenience base properties and realizing base properties is easy enough to grasp. Suppose we grant the general physicalist hypothesis that every mental property is realized by a physical property. Then, for any mental property *M*, each actual or possible physical property which realizes *M* is a realizing base property of *M*. The Boolean disjunction of those realizing base properties is the supervenience base property of *M*.

I will argue that the composite nature of the supervenience base property of any given mental state makes the modal biconditionals licensed by supervenience compatible with the anomalism of mental properties.

Supervenience base (as opposed to realizing base) properties are formed by Boolean disjunction. A major worry with disjunctive properties of the sort envisaged is that it is misleading to call them physical. The point can, perhaps, be best illustrated by way of an example from economics.⁷⁴ Consider, for example, the economic property of being (a piece of) money. Anything could be a piece of money given the right context of scarcity. That is to say, the property of being money can be *variably realized* by different physical properties. The possibility of variable realization, here, is just the possibility that the property of being money is realized, in distinct economic contexts, by distinct physical properties. In the case of money, variable realization does, as a matter of fact, obtain: the property of being money is variably realized by different physical properties in different economic communities. However, it is important to realize that the variable realization of the property of being money is not restricted to the physical properties which do, as a *matter of fact*, realize the property of being money; it also covers all the properties which *could* realize that property. Since anything could count as money given the

74. The Example is borrowed, in a modified form, from Paul Teller, "Comments on Kim's Paper", The Southern Journal of Philosophy vol. XXII Supplement, ed., T. Horgan (The Spindel Conference 1983: Supervenience). The idea derives from Jerry Fodor, "Special Sciences", Synthese, vol.28, pp. 97-115, where it is used for a slightly different purpose.

right context of scarcity, the number of possible realizing properties of money is infinite: the disjunctive supervenience base property which is necessarily co-extensive with the property of being money must be closed under *infinite* disjunction.

The problems with regarding such a disjunctive property as physical stem from the *size* and *heterogeneity* of the property. To call it physical simply because it is a Boolean construction out of physical properties would be misleading. For there is no single physical characteristic which all things satisfying the the individual disjuncts have in common.⁷⁵ Crucially, whatever criteria we have for deciding whether a property is to be included (or excluded) as one of the disjuncts cannot be physical. Any thing could count as money given the right context of scarcity. Therefore, in talking of the *variable realization* of the economic property of being money, the relevant parameter determining how *variable* the *realization* can be is *logical possibility*. Hence, the criteria of inclusion/exclusion cannot be physically circumscribed. The only criteria we seem to possess would be that a given physical property can be included in the disjunctive supervenience base if and only if that physical property *realizes* the property of being money. And this is just to argue that the identity conditions of the disjunctive property (i.e. the supervenience base property) are economic. It is just to argue that the property of being money is an economic property which is instantiated in physical object-tokens.⁷⁶

75. c.f. Teller, "Comments on Kim's Paper", op. cit.

76. c.f. Teller, "Comments on Kim's Paper", op. cit.

Similar remarks apply in the case of psychophysical supervenience. Because of variable realization, the number of properties which may be disjunctively related in the supervenience base of a given mental property is potentially infinite. Moreover, these properties need have nothing *physically* in common with each other. Because variable realization is concerned with *logical* possibility, and because the various disjunctively related properties need have nothing, physically, in common, the criteria according to which we gather the various realizing properties into a single supervenience property cannot be physical. In fact, the only criteria we have for inclusion/exclusion in the supervenience base are mental: *it is a necessary and sufficient condition for membership in the disjunctive supervenience base that the realizing property realizes the supervenient mental property in question.* Specification of the identity conditions of the supervenience base property must make essential reference to the mental property which supervenes on this base. And this is just to argue that the supervenience base property has identity conditions which are mental.

If the above remarks are correct, then strong supervenience involving such supervenience base properties does not undermine the principle of anomalism. To see why, recall the reconstruction of the argument for anomalism expressed earlier in this chapter. The argument was expressed as follows:

- (1) $\square (x)[C_1x \rightarrow n_1x]$
- (2) $\square (x)[n_1x \leftrightarrow m_1x]$
- (3) $\square (x)[C_1x \rightarrow m_1x]$

Given the disjunctive nature of the supervenience base, the argument will become modified to something like the following:

$$(4) \quad \Box(x)[(C_1 \vee C_2 \vee \dots \vee C_n)x \rightarrow (n_1 \vee n_2 \vee \dots \vee n_n)x]$$

And the modal biconditional licensed by supervenience is;

$$(5) \quad \Box(x)[(n_1 \vee n_2 \vee \dots \vee n_n)x \leftrightarrow m_1x]$$

Therefore;

$$(6) \quad \Box(x)[(C_1 \vee C_2 \vee \dots \vee C_n)x \rightarrow m_1x]$$

To properly understand what is going on here, one must recall the earlier distinction between an individuation condition and an identity condition. For present purposes, the distinction comes down to this: an individuation condition V of a property A provides a sufficient condition for the existence of property A , an identity I condition of a property A provides both a necessary and sufficient condition for the existence of A .

In the above schema, $[C_1 \vee \dots \vee C_n]$ is an individuation condition of $[n_1 \vee \dots \vee n_n]$. However, $[C_1 \vee \dots \vee C_n]$ is also a property, and as a property, it has an identity condition. Quite generally, individuation conditions are properties which have identity conditions.

Now, if the above argument is correct, the identity conditions of the disjunctive property $[n_1 \vee \dots \vee n_n]$ are not physical identity

conditions, since their specification must make essential reference to the mental property which supervenes on $[n_1 \vee \dots \vee n_n]$. Therefore, the identity conditions of $[C_1 \vee \dots \vee C_n]$ are not physical. For the individual properties C_1, C_2, \dots, C_n are unified as a single property only insofar as they, together, provide an individuation condition for the disjunctive property $[n_1 \vee \dots \vee n_n]$. Then, if the identity conditions of $[C_1 \vee \dots \vee C_n]$ are not physical, then $[C_1 \vee \dots \vee C_n]$ is not a physical property. But, it is as a property that $[C_1 \vee \dots \vee C_n]$ is an individuation condition. Therefore, $[C_1 \vee \dots \vee C_n]$ is not a physical individuation condition.

Therefore, the above argument schema, (4)-(6), does not jeopardize the principle of anomalism. $[C_1 \vee \dots \vee C_n]$ is not a physical condition of individuation. Thus, m_1 does not receive a physical condition of individuation. Therefore, the fact that supervenience licenses modal biconditionals connecting mental properties with disjunctive supervenience bases is of no consequence for the issue of mental anomalism.

3

The principles enunciated in the preceding section will play a central role in the remainder of this thesis. Consequently, time will now be taken to clarify and defend them. At the same time, the completion of the arguments against psychophysical laws, promised at the end of part 1, will be described.

The major problem with the arguments given in the previous section is that it is not at all clear that the identity conditions of $[n_1 \vee \dots \vee n_n]$ are not physical. Central to the argument that they are not is the claim that specification of these identity conditions must make essential reference to the mental property which supervenes on $[n_1 \vee \dots \vee n_n]$. But it is not at all clear that this fact, in itself, establishes the desired conclusion. Suppose, for simplicity's sake, that each individual disjunct, $n_1, n_2 \dots n_n$ is a physical property. Then, it could be argued, the identity condition of the disjunctive property $[n_1 \vee \dots \vee n_n]$ is the Boolean sum of the identity conditions of each individual disjunct. If so, the identity condition of the disjunctive property appears to be physical.

In response to this objection, the earlier claim has to be modified to the claim that the disjunctive property $[n_1 \vee \dots \vee n_n]$ has no *natural* physical identity condition, hence is not a *natural* physical property. And this requires explanation of the notion of naturalness as understood here.

I will begin with a claim made by David Lewis that natural properties are those which bestow causal powers and which make for resemblance.⁷⁷

77. David Lewis, "New Work For a Theory of Universals", Australasian Journal of Philosophy, vol.61, no.4, pp. 343-77.

Let us begin with *resemblance*. Consider properties such as being green or being square. Typically, attribution of these properties to distinct individuals takes place on the basis of perceived similarities between the individuals, similarities which make for the possession of these properties. That is, attribution of these kinds of properties is grounded in the intuitive belief that the individuals share certain features in virtue of which they possess the common property.

One does not need to be a realist with regard to properties in order to appreciate this intuition. However, these sorts of principles clearly do not seem to apply to the wildly disjunctive properties of the sort which form the supervenience bases of mental properties. In a given disjunctive property, an individual disjunct may have nothing in common with any other disjunct in that property, other than being a member of the same disjunctive type. Hence, two individuals instancing the same disjunctive property may have no feature in common in virtue of which they do so. Hence, disjunctive properties do not make for resemblance in the usual sense of that term.

While it is difficult to deny the intuitive appeal of these remarks, it is also difficult to formulate them in any nonquestion-begging way. In particular, it appears that the notion of resemblance is just as difficult to explicate as the notion of naturalness. Let us, therefore, turn to the second of Lewis' suggestions: natural properties are those which possess causal power.

Lewis, of course, does not distinguish causal power from causal relevance. However, I do not think *my* notion of causal power is sufficient for distinguishing natural from non-natural properties. Consider the wildly disjunctive properties under discussion. Presumably, each disjunct will have causal powers of some sort. So why can we not say that the disjunctive property has causal powers also: these causal powers will be the disjunctive Boolean sum of the causal powers possessed by each individual disjunct. And, indeed, there seem to be no grounds for denying this. Hence, the notion of causal power, as understood in this thesis, does not seem to capture the notion of naturalness.

We can more closely approximate the truth by way of the notion of causal *relevance*. Causal relevance is the property of playing an explanatory role in causal explanatons. The reason is that causal power is *aggregative* in a way that causal relevance is not. The precise claim is this: What makes a property natural is its capacity to figure in explanations, causal or otherwise. That is, *it is the possession by a property of explanatory relevance which will make that property a natural property.*

This account makes naturalness relative to a domain of explanation. Thus, what makes a property a natural *physical* property is its capacity to figure in the (presumably causal) explanations of physical theory (very broadly construed). But what makes a property a natural property of economics, or sociology, or psychology, is its capacity to figure in the explanations of economics, sociology, or psychology respectively.

So what makes a property natural with respect to a domain of explanation d is the possession by that property of explanatory relevance of a sort appropriate to the domain d . Thus, in domains where the only sort of explanation is causal explanation, it will be possession by a property of causal relevance which makes for that property's being natural. And when Davidson argues against psychophysical laws, he is arguing against lawlike connections between mental properties and the properties of a closed comprehensive physical theory. Given Davidson's general attitude toward this sort of theory, it is plausible to suppose that the explanations derived from this theory will be causal explanations. If so, then what makes a property a natural property of the closed comprehensive theory will be its capacity to figure in the causal explanations of that theory. That is, what makes the property natural will be its possession of causal relevance of a sort appropriate to the closed comprehensive physical theory.

In the case where biconditional correlations can be set up between natural properties from two distinct domains of explanation, then the possibility of the reduction, and subsequent identification, of one natural property with the other emerges. However, Davidson's arguments against psychophysical bridge laws are, in the mental-physical case, meant to prohibit such necessary biconditional correlation. Properly understood, then, the argument against psychophysical laws must be formulated as follows:

Psychophysical bridge laws are impossible because such laws would transmit *natural* individuation conditions from the property referred

to by the reducing predicate to the property referred to by the reduced predicate. That is, psychophysical bridge laws are impossible because they would give mental properties *natural* physical conditions of individuation.

Since it is possible to set up modal biconditionals between mental properties and - wildly disjunctive - properties which are in some sense physical, Davidson's argument, if it is to have any plausibility, must be restricted by the *naturalness* requirement. It is only transmission of *natural* physical individuation conditions which is incompatible with mental anomalism.

The disjunctive supervenience base properties, while arguably physical properties, can in no sense be said to be natural physical properties. They certainly figure in none of the explanations of physical theory, whether the theory is of the closed comprehensive variety envisaged by Davidson, or of physical theory as espoused today, or even of folk physical theory, governing relations between ordinary macro-objects.

More fully: The identity condition of the disjunctive supervenience base $[n_1 \vee \dots \vee n_n]$ is not *naturally* physical. Hence, the identity condition of $[C_1 \vee \dots \vee C_n]$ is not *naturally* physical. For the individual properties $C_1 \dots C_n$ are unified as a single property only as they together provide an individuation condition for the disjunctive property $[n_1 \vee \dots \vee n_n]$. Then, if the identity condition of $[C_1 \vee \dots \vee C_n]$ are not naturally physical, then $[C_1 \vee \dots \vee C_n]$ is

not naturally physical; that is, it is not a natural physical property. But, $[C_1 \vee \dots \vee C_n]$, as a property, is an individuation condition. Therefore, $[C_1 \vee \dots \vee C_n]$ is not a *natural* physical individuation condition.

Conversely, the explanations $[n_1 \vee \dots \vee n_m]$ is likely to fit into are explanations of psychology. Due to its necessary biconditional connection with its supervenient mental state, $[n_1 \vee \dots \vee n_m]$ can be substituted into all explanations involving that mental state. Thus, the only *natural* identity condition $[n_1 \vee \dots \vee n_m]$ is likely to possess will be psychological in character. And then the argument can presumably be re-run to show that the only *natural* identity condition $[C_1 \vee \dots \vee C_n]$ is likely to possess will also be mental. The point can be put thus: *any physical specification of $[n_1 \vee \dots \vee n_m]$ or $[C_1 \vee \dots \vee C_n]$ will be patently non-natural, and any natural specification will be patently non-physical.*

Therefore, the argument schema expressed in (4)-(6) does not undermine the principle of the anomalism of the mental. The two are compatible.

4

The above arguments, if correct, show that the modal biconditionals licensed by strong supervenience, which connect mental properties with disjunctively physical supervenience base properties do not undermine

the principle of anomalism. However, the proponent of anomalism still faces a serious problem. Given the strong supervenience of mental on physical properties, it follows that each mental property has a sufficient condition in a property of the supervenience base. That is, every supervenient mental property has a sufficient condition in each realizing base property which is disjunctively related in the supervenience base property. Therefore, the following schema can be set up:

$$(7) \quad \Box(x)[C_1x \rightarrow n_1x]$$

and by supervenience;

$$(8) \quad \Box(x)[n_1x \rightarrow m_1x]$$

Therefore;

$$(9) \quad \Box(x)[C_1x \rightarrow m_1x]$$

where n_1 is a realizing base property of m_1 . Here m_1 does appear to have a physical sufficient condition of individuation. The conditions transmitted by the supervenience conditionals will fail to be completely general individuation conditions: they do not state necessary conditions for the existence of m_1 . But that they be necessary is not required for (7)-(9) to be incompatible with mental anomalism. After all, in the original statement against psychophysical laws based around a modal biconditional connecting n_1 and m_1 , what was incompatible with anomalism was clause (3) $\Box(x)[C_1x \rightarrow m_1x]$. But clause (3) is identical with clause (9). So if (3) is incompatible with anomalism, (9) must surely be also. Therefore, the strong supervenience of mental on physical properties

again seems just as incompatible with anomalism as psychophysical laws would.

The answer to this problem is, I believe, implicit in the answer given to the problem of modal biconditionals. In connection with this, it has been argued that the supervenience base property of any given mental property, that is the property composed of disjunctively related realizing properties, does not have individuation conditions which are both *natural* and *physical*. Any physical individuation conditions it may have are patently non-natural, and the only natural individuation conditions it has are patently non-physical. I think that much the same line of argument can be extended to realizing base properties; these properties do not have natural physical conditions of individuation.

In the discussion of supervenience base properties, it was assumed that these were composed of disjunctively related physical properties. To argue, then, that Boolean disjunctions of these physical properties are not themselves physical properties is to argue that physical properties are not closed under disjunction. And this is precisely what the arguments from variable realization established. I want to claim much the same in the case of realizing properties: realizing base properties are not closed under the modes of construction responsible for them.

The primary construction operation in the formation of realizing properties seems to be *conjunction*. Recall the earlier economic example. Anything could count as a piece of money given the right

context of scarcity. Thus, if economic properties are to supervene on physical properties, the property of being money will have to be understood as a relational property between a physical object-type and the physical environment-type which constitutes the economic context.⁷⁸ Consequently, specification of any realizing property of money will, necessarily, have to include a physical description of the whole economic context. Therefore, the disjunction which provides a necessary condition for the property of being money will have to include full physical descriptions of all economic systems, and these will be of the following form: "... is such and such a piece of copper in such and such a physical environment, or in such and such a physical environment ... or is a piece of shell in ... ". The realizing property is correctly to be understood as a conjunction of physical object-type and physical environment-type. Therefore, specification of any realizing property (at least for economic properties) makes essential use of *conjunction* as a property-forming operation.

These remarks seem to apply quite clearly in the case of mental properties also. At least they apply quite clearly in the case of mental properties which, in contemporary literature, have become known as *wide* mental states. The notion of a wide mental state was characterized earlier. Most folk psychological states are wide in that they are composed of both internal and external components. The

78. This point is also made by Teller, "Comments on Kim's Paper", *op. cit.*, Teller, however, does tend to conflate the question of whether a supervenience base is disjunctive from the question of whether it is conjunctive. One of the principal reasons for distinguishing supervenience and realizing bases is to avoid such conflation.

important characteristic of wide mental states, for present purposes, is that they are essentially *world-involving*. A precondition of having the belief that snow is white, for example, is that the subject of the belief has, at some time, been in some appropriate causal or environmental relation to snow, and also to instantiations of the property of whiteness. Now consider the ramifications this has for the nature of the realizing base property of that mental state. Because the belief is, in the above sense, externally individuated, the property which forms the realizing base of the belief must be understood as a relational property between some appropriate neurophysiological state occurring 'in the head', appropriate external state of affairs in the world (i.e. an arrangement of objects and properties in the world), and appropriate connecting relation. Thus conjunction is an essential property-forming operation in the construction of realizing base properties of wide mental states.

Now it is important to realize that for an argument of the form expressed in (6)-(9) to undermine the principle of anomalism, the realizing base property n_1 must have conditions of individuation C_1 which are natural and physical in character. But, as has been argued above, the realizing base property of an ordinary wide mental state will be a complex property constructed out of simpler properties by the operation of conjunction. Let us suppose that these simpler properties - the neurophysiological state, the state of affairs, and the connecting relation - are all physical properties. Enough has been said by now to undermine confidence in the assumption that the constructed property is also natural and physical. But only on this assumption will the

argument expressed in (6)-(9) undermine the principle of anomalism. I shall argue that this assumption is untenable. This will be the subject of the following chapter.

The remainder of this thesis can be seen as an extended discussion of the principles which have emerged in the last few sections. That is, the remainder of this thesis can be seen as a specific application, firstly to mental states, then to syntactic properties, of the following principle:

- (α) If property M is (a) anomalous, and (b) is strongly supervenient on supervenience base property S composed of realizing base properties $R_1 \dots R_n$ then (i) S has no natural specification independently of M , and (ii) none of $R_1 \dots R_n$ have a natural specification independent of M .

**Anomalism, Supervenience, and the Dual Component Theory
of Mental Properties**

1

This Chapter will continue discussion of the problem of reconciling the principle of the anomalism of the mental with the claim that mental properties are strongly supervenient on physical properties. The principle concern of this Chapter will be the problems posed for this reconciliation by the following schema:

$$\begin{aligned}
 (R) \quad & \Box (x)[C_1 x \rightarrow n_1 x] \\
 & \Box (x)[n_1 x \rightarrow m_1 x] \\
 & \Box (x)[C_1 x \rightarrow m_1 x]
 \end{aligned}$$

where, m_1 is a supervenient mental property, n_1 is a realizing base, and C_1 is an individuation condition of n_1 .

In Chapter III, it was suggested that a possibility for making this schema compatible with the anomalism of m_1 , lies in the *composite* nature of n_1 , and, consequently, of C_1 . The principle underlying this suggestion is that principle elicited in the earlier discussion of modal biconditionals licensed by strong supervenience. The idea is that the composite nature of n_1 and, hence, of C_1 , precludes either possessing

identity conditions which are both physical and natural. It will be argued that this is indeed the case.

The compositeness of n_1 and C_1 stem from the fact that m_1 is itself composite. And m_1 is composite because, as a mental state, it is externally individuated and, so, world involving. Accordingly, discussion of the above schema, and the possibilities for its reconciliation with mental anomalism will take place within the framework of the dual component theory of mental properties.

2

The argument schema represented in (R) creates the following problems for the principle of the anomalism of the mental. The modal supevenience conditionals transmit conditions of individuation from the realizing base property n_1 to the mental property m_1 , giving m_1 physical - and, hence, nonrational - conditions of individuation. And this would break the connection between m_1 and the principles of rationality to which m_1 is essentially, or constitutively, subject. It has been argued that for such consequences to ensue, the transmitted conditions of individuation must be both physical *and* natural.

The first, and extremely important, point to note is that m_1 is a *wide* mental state - externally individuated and, hence, world involving. And this holds quite generally. Although Davidson's discussion of these matters does not operate within a dual component account of mental

states, it is quite clear that the states which are essentially subject to the norms of rationality are mental states widely individuated. A corollary of this is that if the transmission of natural, physical (hence, nonrational) individuation conditions is to pose a threat to the constitutive status of rationality, and thus to the principle of anomalism, these individuation conditions must be transmitted to *wide* mental states.

Consider, now, the mental state *narrowly* individuated. The narrow mental state, or the *internal component*, is the locus of causal power. In addition, if an 'explanatorily narrow' interpretation of the dual component theory is adopted, the narrow mental state is also the locus of causal (hence, explanatorily) relevance. However, what cannot be said, yet, is that narrow mental states are constitutively governed by principles of rationality. That is, the inference from the claim that wide mental states are essentially subject to norms of rationality to the claim that narrow mental states are essentially governed by those norms is not, in general, valid, even though narrow mental states are proper parts of wide mental states. It may be that wide mental states are essentially subject to norms of rationality because of some component of their content which falls outside the sphere of narrowness. That is, it may be something about the external component of the wide mental state which makes that state essentially subject to norms of rationality.

However, this response does seem to possess a certain counterintuitiveness, perhaps best expressed in the interrogative: given

that mental states are essentially subject to rationality, and given that narrow mental states, or internal components, form parts of mental states, how can they avoid being essentially subject to rationality also?

The answer is that internal components can avoid being *essentially* subject to principles of rationality only if they are not an *essential* part of mental states. For, if this were the case, it would be possible for internal components to be subject to principles of rationality without being *essentially* subject to those principles. That is, internal components would be subject to principles of rationality because they are proper parts of mental states, but they would not be essentially subject to those principles because they are not essential parts of mental states.

This position can only be maintained if one adopts what was earlier characterized as *total externalism** with regard to content. The total externalist must hold that mental states are essentially subject to the norms of rationality in virtue of their external component. The reason is not that total externalists* have no room in their theory for the notion of an internal component, or narrow mental state; they need not deny that mental states have properties constitutive of an intra-individual causal role. The problem is rather, this: According to total externalism*, the properties constitutive of an intra-individual causal role play no part in the type-individuation of a mental state. And if they play no part in its type-individuation, they cannot be *essential* to that mental state. Therefore, mental states, on the total

externalist hypothesis, cannot be subject to the norms of rationality in virtue of their internal components. For mental states are essentially subject to the norms of rationality, and they cannot be essentially subject to anything in virtue of properties they have only contingently (viz. properties of intra-individual causal role). Therefore, if total externalism* is true, mental states must be subject to norms of rationality in virtue of their external components, for, in this case, only the external component is had essentially by any given mental state. Moreover, only if one adopts total externalism* does it seem possible to motivate the view that rationality applies to mental states purely in virtue of their external components. For, once one allows that internal components are a type-individuating, and hence essential, part of wide mental states, there seem to be no grounds for denying that internal components are also essentially subject to norms of rationality.

Given the correctness of these remarks, a reconciliation of the principles of anomalism and supervenience can now be effected:

Firstly, suppose a principle of total externalism* with respect to the individuation of mental states is adopted: the only type individuating feature of a mental state is its external component. As was argued above, a corollary of this claim is that rationality constitutively governs relations between mental states in virtue of the external component of those states. In particular, relations between internal components, or narrow mental states, are not *essentially* subject to norms of rationality. However, if this is so, then the receiving by narrow mental states of individuation conditions which are

natural, physical, and, hence, nonrational, would not undermine the constitutive status of rationality, hence, would not undermine the anomalism of mental properties. For relations between narrow mental states are not constitutively governed by principles of rationality.

Therefore, it is possible to accept, consistently with acceptance of the anomalism of the mental, that narrow mental states strongly supervene on physical states in such a way as to receive individuation conditions which are both physical and natural. The consistency of the two claims is, in this case, made possible by acceptance of the principle that wide mental states are essentially subject to norms of rationality in virtue of their external components. And this claim is, in turn, made possible by the total externalist* hypothesis that the only type-individuating feature of a mental state is its external component.

It is possible to hold, therefore, that mental properties are simultaneously anomalous and strongly supervenient on natural physical properties: mental properties are anomalous in virtue of their external component, and they are strongly supervenient on natural physical properties in virtue of their internal component. The point can be put in a way less likely to engender confusion as follows. Mental properties are themselves possessors of (higher-order) properties. Some of these properties - those constitutive of the external component - are anomalous; they can enter into no modal correlations with natural physical properties, whether these correlations are of a conditional or biconditional character. However, others of these properties - those

constitutive of the internal component - can enter into modal conditional correlations with natural physical properties.

On this view, then, internal components of mental states receive natural physical conditions of individuation. And then the possibility of a nomothetic science of psychology opens up. That is, the possibility emerges of a cognitive psychology conceived of as detailing interactions between internal components or narrow mental states. These are respecters of (explanatory) methodological solipsism, and they derive their explanatory relevance from the underlying natural physical properties which realize them. This matter will be discussed later.¹.

It is interesting to note that, at least in his later writings on content, Davidson can be plausibly construed as adopting a total externalism* with regard to mental content. Consider, for example, the following passage:

" My main point is that our basic methodology for interpreting the words of others necessarily makes it the case that most of the time the simplest sentences which speakers hold true are true. It is not the speaker who must perform the impossible feat of comparing his belief with reality; it is the interpreter who must take into account the causal interaction between world and speaker in order to find out what the speaker means, and hence what he believes the sentences that express the beliefs, and the beliefs themselves, are correctly understood to be about the public things and events that cause them, and so must be mainly veridical."²

1. In particular, see Appendix III.

2. Donald Davidson: "Empirical Content", in E. LePore ed., Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson (Oxford, Blackwell, 1986), pp. 320-332, p. 332.

If passages such as this are indicative of a total externalism*, Davidson has a way of reconciling the principle of the anomalism of the mental with the strong supervenience of mental on physical properties. What is perhaps surprising is that an adequate defence of Davidson's views on psychophysical laws, and hence of his token identity thesis, requires essential use of his later views on content-individuation. Davidson nowhere makes this claim.

3

The reconciliation outlined in the previous section is essentially incomplete. It will want to be maintained, as part of a general physicalist thesis, that, for any given mental state, all of it - i.e. both internal and external components - strongly supervenes on a physical property or state. Can this be consistently maintained together with the claim that mental states are anomalous? I will argue that it can. The reason relates back to the discussion in Chapter III concerning the naturalness of supervenience and realizing bases.

It was argued in Chapter III that for the *prima facie* incompatibility of anomalism with strong supervenience to be a real incompatibility, the identity conditions of supervenience and realizing bases had to be natural in the sense characterized in that Chapter. It will be argued that the strong supervenience of wide mental states on physical properties is compatible with the anomalism of wide mental states because the realizing bases of such states do not constitute

natural physical properties, whose identity conditions can be specified in natural physical terms.

According to the model now under discussion, the internal component of a mental state is realized by a state of the brain (or, perhaps, of the whole organism). We may assume - and the model requires this assumption - that such a state has identity conditions which can be specified naturally in terms which are purely physical. In contrast, a mental state widely individuated will not be realized by a state of the brain (or organism). Rather, it will be realized by a combination of brain state, state of affairs (that is, an arrangement of objects and properties in the world), and a relation between them. I will argue that the type of realization implicated here does not jeopardize the principle of anomalism because such a combination of brain state, state of affairs, and connecting relation, does not constitute a distinct natural property; that is, it does not constitute a property whose identity condition can be specified in natural physical terms.

The reason is that the only (natural) criterion we have for regarding the combination of brain state, state of affairs, and connecting relation as a *single* property is that they together realize the mental state which supervenes on them. That is, a sufficient condition for conjoining the brain state, state of affairs, and connecting relation, into a single property is that they *together* realize the appropriate mental state. Consequently, (natural) specification of the identity condition of the realizing base property makes essential use of mental terms. Therefore, the realization of a

wide mental state by such a base does not entail that the mental state receives natural physical conditions of individuation. Whether or not the combination of brain state, state of affairs, and connecting relation possesses an identity condition specifiable in purely physical terms, any such specification cannot be natural.

More fully: It is not sufficient for the existence of a realizing base property of a wide mental state that there exist a certain brain state b , and a certain state of affairs f . For sufficiency, there must also be an appropriate relation R between b and f . What is this relation R ? Well it seems the only way we can specify what R is, in anything like natural terms, is by saying something like " R is the relation such that when bRf then M ". That is, R is the relation such that when b bears R to f they together realize the wide mental state M . Thus, reference to mental state M is an essential part of any natural specification of R , and this means that a natural specification of the identity condition of the compound state $[bRf]$ makes essential reference to mental state M . Therefore, the identity condition of the realizing base property $[bRf]$ is not a natural physical identity condition: the only reason we have for regarding $[bRf]$ as a single property is that, together, b , R , and f realize M . Therefore, $[bRf]$ is not a natural physical property. Therefore, the strong supervenience of wide mental state M on the state $[bRf]$ does not entail that M receives a natural physical individuation condition.

This may seem counterintuitive. If we suppose that the brain state, the worldly state of affairs, and connecting relation have,

individually, identity conditions which are natural and physical, then the identity condition of the compound property [bRf] can be specified by conjoining the identity conditions of the three components. The crucial point, however, is that while it might be possible to arrive at an identity condition by such means, such a condition could not be natural. The property [bRf] is not a natural physical property: it plays no role in any explanation of any physical theory. Hence, the considerations which would lead us to unify b, R, and f into a single property cannot be considerations derived from physical theory. And this is what it means to say that [bRf] is not a natural physical property.

This argument is a specific application of the discussion of anomalism, supervenience, and naturalness developed in Chapter III. In that Chapter, it was argued that strong supervenience is incompatible with anomalism only if it entails transmission of natural physical conditions of individuation. And for such transmission to occur, the identity conditions of the supervenience or realizing base properties (depending on the type of correlation in question) must be natural and physical.

The composite nature of the realizing base property of the wide mental state M requires the following modification to the argument schema represented in (R):

$$(R^*) \quad \square (x)[(C_{1b} \ \& \ C_{1R} \ \& \ C_{1f})x \ \rightarrow \ (n_{1b} \ \& \ n_{1R} \ \& \ n_{1f})x]$$

$$\square (x)[(n_{1b} \ \& \ n_{1R} \ \& \ n_{1f})x \ \rightarrow \ Mx]$$

$$\square (x)[(C_{1B} \& C_{1R} \& C_{1f})x \rightarrow Mx]$$

Where n_{1B} is the brain state, n_{1f} is the state of affairs in the world, and n_{1R} is the connecting relation. C_{1B} , C_{1f} , and C_{1R} , are the respective individuation conditions of these properties.

It is because $(n_{1B} \& n_{1R} \& n_{1f})$ has no natural physical identity condition that $(C_{1B} \& C_{1R} \& C_{1f})$ has no natural physical identity condition. Hence, $(C_{1B} \& C_{1R} \& C_{1f})$ is not a natural physical property. But as a property, $(C_{1B} \& C_{1R} \& C_{1f})$ is an individuation condition. Therefore, $(C_{1B} \& C_{1R} \& C_{1f})$ is not a natural physical individuation condition. Therefore, the schema (R^*) does not entail that M receives a natural physical individuation condition. The principles elicited here are familiar from Chapter III.

It will be instructive here to compare the psychophysical case with other examples of supervenience. The case of economic properties has been discussed in Chapter III. There, it was pointed out that economic properties are to be correctly understood as relational properties, relations between physical object-types and physical environment-types. As such, they share with mental properties all the relevant characteristics of external individuation and world-involvingness necessary to re-run the argument described above. The realizing bases of economic properties cannot be plausibly construed as natural physical properties. Hence, the strong supervenience of economic properties on these bases would not involve the transmission of natural physical conditions of individuation.

Another case in point is the supervenience of *moral* on physical properties. Suppose we accept the claim that moral properties supervene on physical properties. Then, given an action to which we attribute the property of being good, we would want to say that any physically relevantly similar action in any physically relevantly similar circumstances, would also be good. Now consider such an action. For example, swimming a river to save a drowning child. Suppose we say that this action is good. Given the supervenience of moral on physical properties, we would be committed to saying that any physically relevantly similar action in physically relevantly similar circumstances would also be good.

The physical properties (instantiated) which form the action characterized as 'swimming a river to save a drowning child', then, constitute a sufficient condition for the moral property of being good; they form the realizing base property of the property of being good. However, this base must include at least the following. Firstly, certain movements of the arms, legs, torso etc. which could constitute the person's swimming. Secondly, these movements must take place in an appropriate milieu i.e. liquid - if there was no liquid present, the person could not be swimming, nor the child drowning. Thirdly, there must be a child involved, otherwise the action could not be correctly characterized as 'saving a drowning child'. So the realizing base property must be regarded as a conglomeration of bodily movements directed toward child in appropriate milieu. It is fairly clear that such a conglomeration does not form a natural physical property. The property surely cannot play a role in any of the explanations of

physical theory, it does not have explanatory relevance of a sort appropriate to physical theory. And the identity condition of the property seems to be derived more from considerations of action theory: the individual properties are unified into a single property upon the condition that they together constitute the action characterized as swimming a river to save a drowning child. The supervenience of moral properties is, then, in this regard, relevantly similar to the supervenience of wide mental states.

4

With regard to the particular reconciliation of anomalism and supervenience proposed in these sections, the following overall picture emerges. Internal components, or narrow mental states, are strongly supervenient on natural physical properties. They receive natural physical conditions of individuation and, hence, are not anomalous. This leads to the possibility of a nomothetic science of psychology conceived of as detailing interactions between internal components of mental states, where these components obtain their explanatory relevance from the natural physical properties which provide their conditions of individuation. Wide mental states, in contrast, are strongly supervenient upon states which are in some sense physical, but these states are not natural physical states. This is compatible with the anomalism of wide mental states because the transmitted individuation conditions implicated by strong supervenience are not natural physical individuation conditions.

This picture presents an plausible and, in many senses, an appealing reconciliation of the principles of anomalism and supervenience as they apply to mental properties. However, in Part 2, it will be argued that this particular attempt at reconciliation must be rejected.

Narrow Mental States and Rationality1

The reconciliation of anomalism and supervenience, as applied to mental states, advanced in part 1 allows internal components or narrow mental states to strongly supervene on natural physical states, and, thus, to receive natural physical conditions of individuation. In part 2, it will be argued that this picture cannot be sustained. Narrow mental states cannot be viewed as possessors of natural physical individuation conditions.

The reconciliation of anomalism and supervenience described in part 1 proceeds on the assumption of the principle of total externalism*: the only type-individuating feature of mental states is the external component. For only upon this assumption is it possible to motivate the following claim:

(ER) For any given (propositional) mental state, rationality, as a constitutive principle, applies to that state purely in virtue of its external component.

In deference to the ambiguity of the 'in virtue of' clause, the External Rationality (ER) claim will be understood as the claim that the canons

of rationality apply to the external components of mental states, and then, derivatively, to mental states themselves. It will be argued that the ER claim must be rejected, and the consequences of this rejection will be described.

In this thesis, the rationality of an organism will be assimilated to the conformance of that organism to rules or principles of logic. More precisely, canons stating what relations can correctly be regarded as rational relations will be identified with those rules and principles delineating what relations can correctly be regarded as logical relations. 'Logic' here is to be construed broadly, to be viewed as consisting of classical deductive logic, inductive logic, and, perhaps, abductive logic.

A few points of justification may be in order. The primary purpose of this assimilation is to give a fairly precise sense to what has become a loosely defined notion. Moreover, whether or not there exist principles of rationality constitutively governing mental state relations which cannot be captured under the 'logic' rubric, it is fairly clear that the principles of deductive, inductive, and abductive logic capture many, and almost certainly most, of the principles which are essential to the notion of rationality.

The claim to be discussed, the ER claim, is that rationality applies to, and only to, the external component of mental states. This idea needs to be cashed out in the following way. Rationality, understood as above, applies to, and holds between, that component of

content which constitutes the external component of mental states. This component of content is itself constituted by referential truth-conditional properties. And so a natural way to explicate the ER principle would be in terms of the relation between logic, or conformity to logic, and referential truth-conditions. And the primary relation, here, is one of *interpretation*.

Consider classical deductive logic, restricted, for the sake of simplicity, to the propositional calculus. Firstly, consider this logic from a purely formal or syntactic viewpoint, that is, viewed as a formal uninterpreted grammar. The language will consist of various propositional variables P, Q, R... and various logical connectives, -, &, v, \rightarrow . Then it is possible to give this uninterpreted calculus a truth-theoretic semantics, a semantics based around the notion of truth, as follows.

Firstly, for each propositional variable we can give a truth-definition, something along the following lines:

'P' is true if and only if P

The sentence 'P' which expresses the proposition that-P is true if and only if P. Given simple truth-definitions such as that given above, it is possible to build up more complex truth-definitions - truth-definitions of complex sentences, built up from simple sentences by way of the logical connectives.

Moreover, given these truth-definitions and the notion of truth-preservation, it is possible to characterize a notion of logical

consequence; hence, a notion of validity: A sentence B is a logical consequence of a set of sentences $A_1, A_2, A_3 \dots$ if and only if the inference from $A_1, A_2, A_3 \dots$ to B necessarily preserves truth. In more detail: A sentence B is a logical consequence of a set of sentences $A_1, A_2, A_3 \dots$ if and only if, in the sequent expression $A_1, A_2, A_3 \dots \vdash B$, for every assignment of the value the True (T) to the sentences $A_1, A_2, A_3 \dots$ then, necessarily, B also takes the value the true. Equivalently, if it is impossible to assign the value T to all of $A_1, A_2, A_3 \dots$ and the value F to B. (The sentences $A_1, A_2, A_3 \dots$ and B may either be expressions of simple propositional variables, or constructions out of these).

For example, let the expression $A_1, A_2 \vdash B$ denote the following sequent expression:

$$P \rightarrow Q, \neg Q \vdash \neg P$$

In this sequent expression, $\neg P$ is a logical consequence of the two sentences $P \rightarrow Q$, and $\neg Q$. And, on a truth-theoretic interpretation, this means simply that, necessarily, in every interpretation where $P \rightarrow Q$, and $\neg Q$ take the value T, then so does $\neg P$. Therefore, the inference from the two sentences to the conclusion is said to be a valid inference.

The salient point is this. The connection between the rules of logic and the external component of a mental state is that the referential truth-conditional properties constitutive of the external component can provide an interpretation of those rules. This

interpretation is based around the notion of truth: logical rules or laws are those which necessarily preserve truth. It will be argued that rules of logic, given this sort of truth-theoretic interpretation, cannot function in a way we intuitively require of a principle constitutively governing relations between mental states.

2

The assimilation of canons of rationality to rules of logic, where this logic is given a truth-theoretic interpretation entails the assimilation of rational relations to those relations which necessarily preserve truth. This provides the first major source of unease about the assimilation. For it entails that rationality be assimilated to deductive logic alone. The principal feature of inductive and abductive inferences is that they do not purport to necessarily preserve truth, but only to *probably* do so. However, it is surely an intuitive desideratum of any adequate account of rationality that some accommodation should be made in that account for principles governing inductive generalization on the basis of evidence, and governing abductive inference to the formation of hypotheses. Intuitively, such constraints seem to fall under the domain of rationality.

There are two problems here which need to be distinguished. The first is a problem of scope: The truth-theoretic interpretation of logic, hence of rationality, does not cover all instances of relations which we would intuitively want to call rational. That is, that a

relation be a necessarily truth-preserving relation is not *necessary* for it to be a rational relation. It could also be a rational relation in virtue of being a probably truth-preserving relation. The other problem is one of sufficiency: that a relation be a necessarily truth-preserving one is not *sufficient* for it to be a rational relation. It will be argued that the notion of rationality involves more than just the notion of necessary truth-preservation, it also involves, principally, considerations of *evidence*. And this would be to see *all* rational relations as *probably* truth-preserving. The condition of sufficiency will be concentrated upon here, as this more adequately gets to the core of the problem with a truth-theoretic interpretation of rationality.

Consider, firstly, a *bad* argument for the conclusion for the conclusion that the truth-theoretic notion of rationality is inadequate as a characterization of the principle which constitutively governs relations between mental states. The argument goes something like this:³

If the laws of logic are given a truth-theoretic interpretation, then statements of logical relations are deductively closed under their entailments. Therefore, if canons of rationality are to be assimilated to rules of a truth-theoretic logic, and, hence, if truth-theoretic

3. It is difficult to find a published source for this argument. However, it is possible to interpret Jerry Fodor's "Three Cheers For Propositional Attitudes", in his Representations (Massachusetts, MIT Press, 1981), pp. 100-23, as giving essentially the same argument. As will become clear, I think that such an interpretation of Fodor would be wrong.

logical rules constitutively govern relations between mental states, beliefs (taken here as paradigmatic cases of propositional mental states) would be closed under their entailments. But if this is so, then, for any belief, we would be justified substituting in the sentence which expresses that belief's content, co-referring terms and expressions. That is, we would be justified in reading belief statements transparently, or *extensionally* rather than opaquely, or *intensionally*. But, so the argument goes, the intensionality of belief-contexts is a central characteristic feature of the mental, a feature which any theory of the mental, including a theory about the principles constitutively governing mental state relations, should respect. Hence, rationality cannot be assimilated to logic, where this is given a truth-theoretic interpretation.

The reason this is a bad argument is that it confuses *normative* and *descriptive* functions of rationality. This subject was touched upon in Chapter III, and will be extensively discussed later. The basic idea is that in all talk of rationality being constitutive of the mental, the notion of rationality at work is a normative rather than a descriptive notion. That is, rationality is properly taken to be an *ideal* which actual organisms only approximate. Therefore, it would not do to object to the truth-theoretic interpretation of rationality on the ground that no-one actually believes all the consequences of their beliefs, and hence that beliefs are not, actually, deductively closed under their entailments. This would be to confuse normative and descriptive notions of rationality. In order to show that the truth-theoretic account provides an inadequate characterization of the concept of rationality,

it is necessary to show the inadequacy of the characterization when rationality is conceived normatively.

A beginning can be made at strengthening the above argument in the suggested direction by noting that there are at least some cases where we do not assume the simple deductive closure of beliefs under entailment *even as a normative principle*. The point has been extensively discussed by Fodor.⁴ Consider Dennett's, now famous, example of the chess playing machine.⁵ Dennett uses this to argue for the rationality maximization principle. Fodor sees himself as using it to establish the opposite conclusion:

".... the hope one cherishes in playing chess with a machine (or with, for that matter, some better instance of an intentional system - say, a Russian) is that (s)he/it *won't* make the best, most rational, optimal (etc.) move, but will, instead, make the very move that falls for one's little ruses, springs one's little traps, and, in general, exhibits levels of intellectual functioning gratifyingly inferior to one's own. Trap-baiting is itself rational only in the light of such hopes and predictions it is precisely from the intentional stance that we hope and fondly expect that our opponent will not *notice* our traps, or that if our traps are noticed they will not be *understood*, or, if they are understood, no way will they be *seen, remembered, or otherwise conjured up*".⁶

The first point to note is that while Fodor sees the argument as establishing that the rationality maximization principle cannot be sustained, the most the argument can be said to establish is that the rationality maximization assumption must be rejected *if that rationality*

4. "Three Cheers For Propositional Attitudes", *op. cit.*

5. "Intentional Systems", *op. cit.*

6. "Three Cheers For Propositional Attitudes", *op. cit.* p. 106.

is given a truth-theoretic interpretation, that is, if rational relations are assimilated to logical relations where that logic is given a truth-theoretic interpretation.

How, exactly, does it establish this conclusion? One might seek to put the point thus: To the extent that we assume the deductive closure under entailment as a normative principle, or as an ideal, then, to that extent, we are also justified in assuming the substitution of co-referring in the content-sentences of beliefs as a normative principle. That is, to the extent we assume the deductive closure of beliefs as an ideal, then, to that extent, we assume the extensionality of belief-contexts as an ideal. And this ideal seems misplaced.

The problem with this argument is that it is not at all clear that we do not assume the extensionality of belief-contexts, and the consequent substitution of co-referring expressions, as an ideal. That is, it seems perfectly coherent, indeed plausible, to regard extensionality of belief-contexts as an ideal which is never realized in actual organisms due, perhaps, to the finite character and, hence, partial perspectives of those organisms.

A much better way of putting the point would be as follows. Even if rationality is conceived of as a normative principle, any assumption of deductive closure of beliefs, and, hence, of the extensionality of belief-contexts, is necessarily constrained by considerations of *evidence*. Baldly put, we expect an organism's beliefs to be deductively closed *to the extent that the organism has available to it (or we*

believe has available to it) evidence, on the basis of which it would be (or we believe it would be) rational to infer a given conclusion. The ideal case might be the limiting case where the (ideal) organism's accessibility to the evidence is itself ideal. But this does not mean, of course, that considerations of evidence play no essential constraining role. It is just that, in this case, the evidence that the organism possesses is *all* the evidence there is.

Therefore, the principle we assume to have normative status with regard to the relations between mental states is not *deductive closure of beliefs*, but rather, *deductive closure of beliefs if the evidence is willing*. Therefore, the concept of rationality, conceived of as a normative principle governing relations between mental properties, cannot be assimilated to logic under a truth-theoretic interpretation. Any truth-theoretic characterization of the notion of rationality must be constrained, and, hence, modified, by considerations of evidence. I will call this the *rationality-in-the-face-of-the-evidence* requirement.

The rejection of the ER principle follows by modus tollens. If the ER principle is true, and rationality applies to mental states purely in virtue of their external components, then the rationality which governs mental state relations must be truth-theoretic in character. However, as the above arguments are intended to show, the rationality which normatively governs mental state relations cannot be of a truth-theoretic character. Therefore, rationality cannot apply to mental states purely in virtue of their external components. The ER principle must be rejected.

Before concluding this Chapter, it will be interesting to see if a characterization of rationality can be developed which can adequately play the role of mental constitutive principle. A desideratum of such an account would be that it accord with the rationality-in-the-face-of-the-evidence requirement, and accomodate the essential role played by the notion of evidence. Such a conception can be extrapolated from the work of Hartry Field.⁷ Field has argued that it is possible to give classical deductive logic an interpretation based around the purely epistemic notion of subjective probability. This would yield the required notion of rationality-in-the-face-of-the-evidence. This is not the place for detailed examination of Field's view. My purpose, rather, is merely to sketch the outlines of his approach, and to show how an adequate characterization of rationality can be derived from principles of logic if that logic is given an epistemic rather than a truth-theoretic interpretation.

The process of giving a truth-theoretic interpretaion to a formal uninterpreted grammar has already been looked at. Given various propositional variables P, Q, R, \dots and given various logical connectives, $\neg, \&, \vee, \rightarrow$ one first proceeds to give the variables truth-definitions. These truth-definitions provide the semantics for, initially, the simple propositional variables, then, derivatively, for complex sentences

7. "Logic, Meaning, and Conceptual Role", The Journal of Philosophy, vol. LXXIV, pp. 379-409.

constructed out of the individual variables, and finally for the connectives. Using these truth-definitions and a notion of truth-preservation, it is possible to characterize a notion of *logical consequence*, hence a notion of validity. A sentence B is a logical consequence of a set of sentences $A_1, A_2, A_3 \dots$ if and only if the inference from $A_1, A_2, A_3 \dots$ to B necessarily preserves truth.

The other half of characterizing a formal language is proof theory. Here one provides definitions of proof by appeal to axiom schemata and rules of inference. The important notion of *deducibility* is defined in terms of the existence of a proof.

It is important at this level to distinguish the notion of *logical consequence* from the notion of *deducibility*. The notion of logical consequence is defined in terms of the notion of truth. Hence, it is a semantic notion. In contrast, the notion of deducibility is defined in terms of the existence of a proof. Logical proofs depend on the form or syntactic structure of the involved sentences. Hence, the notion of deducibility is a syntactic notion.

Given a semantics and a system of proof, it is then possible to prove soundness and completeness theorems. The proof system is sound if every sentence deducible from a set of sentences is a logical consequence thereof. The proof system is complete if every logical consequence of a set of premisses is deducible therefrom. Note the interlocking role the notions of deducibility and logical consequence play in these definitions.

Field's project in "Logic, Meaning, and Conceptual Role" is to show that classical deductive logic can be given an interpretation or semantics based not on the notion of truth, but rather on the notion of conditional subjective probability such that, with respect to this semantics, the logic can be given soundness and completeness proofs.

In dealing with classical deductive logic, Field is, of course, dealing with the same uninterpreted formal grammar; he is dealing with the same syntax. Therefore, he can make use of the same proof theory, and can employ the same notion of deducibility, since this is a syntactic notion. However, demonstration of soundness and completeness proofs must also make essential use of the notion of logical consequence, and this is a semantic notion. Demonstration of soundness and completeness, that is, is always with respect to a semantics.

If logic is given a truth-theoretic semantics, the notion of logical consequence is characterized on the basis of truth-definitions and the notion of truth-preservation. Therefore, when Field attempts to give logic a probabilistic semantics, part of his job, at least, will be to show how the notion of subjective conditional probability is can be used to adequately characterize the notion of logical consequence.

Field takes his departure from Popper's axiomatization of conditional probability.⁸ The axiom system provided by Popper is as follows:

8. "Logic, Meaning, and Conceptual Role", op. cit., pp. 381-82.

- I. $P(A/B) \geq 0$
- II. $P(A/A) = 1$
- III. $\exists A \exists B (P(A/B) \neq 1)$
- IV. If $P(A/B) = 1$ and $P(B/A) = 1$
then $\forall C [P(C/A) = P(C/B)]$
- V. $P(A \wedge B/C) = P(A/B \wedge C) \cdot P(B/C)$
- VI. $P(A \wedge B/C) \leq P(A/C)$
- VII. $P(\neg B/A) = 1 - P(B/A)$ unless $\forall C [P(C/A) = 1]$

(where, for example, $P(A/B)$ is read, 'the probability of A given B').⁹

Any function which satisfies axioms I-VII Field calls a *conditional probability function*. Using I-VII we can define the following:

A is *certain* means $\forall C [P(A/C) = 1]$.

A *legitimizes* B (represented $A \leq B$) means $\forall C [P(A/C) \leq P(B/C)]$.

A is *equipollent* to B (represented $A \sim B$) means A legitimizes B and B legitimizes A.

Popper uses the apparatus of conditional probability as a way of *extending* classical logic. For example, legitimization is to be viewed as an extension of the notion of entailment. It is a weaker relation than entailment, but still a logical relation, in some broader sense of 'logical'.

9. The notion employed here, then, is *conditional*, rather than simple, probability. "Logic, Meaning and Conceptual Role", op. cit., p. 381.

Field puts the axiom system to a different use. Instead of viewing the probabilistic apparatus as a means of extending or generalizing classical logic, in Field's work they are viewed as a means of providing an alternative semantics for the logic; a semantics which is not truth-theoretic.¹⁰ The following remarks are meant to constitute merely the basic outline of Field's procedure.

Let us begin with the formal uninterpreted grammar which provides the syntax for classical propositional logic. My remarks will, for simplicity's sake, largely be confined to propositional languages. The first stage in providing a truth-theoretic semantics for this propositional syntax will be the assignment to each propositional variable of a truth-definition. Corresponding to this in the probabilistic interpretation will be the assignment to each variable of a subjective conditional probability function saying how a subject's dispositions to assign probability values to a given variable will depend upon his dispositions to assign probability values to other variables.

The next stage in the truth-theoretic interpretation is to combine the truth-definitions with the notion of truth-preservation in order to characterize a notion of logical consequence. Logical consequence is defined in terms of the necessary preservation of truth. The probabilistic analogue of truth-preservation is preservation of

10. "Logic, Meaning, and Conceptual Role", op. cit., pp. 382-83.

certainty. And preservation of certainty can apply to sentences and inferences.

Consider, first, sentences. Take any sentence which expresses a logical law. For example:

$$[(P \rightarrow Q) \rightarrow (-Q \rightarrow -P)] \quad (A)$$

On a truth-theoretic interpretation this sentence is a logical law because it is true under all interpretations of its propositional variables. That is, it is true under all assignments of truth-values to its variables. That is to say, it is necessarily true.

It is possible to characterize the probabilistic analogue of necessary truth in terms of certainty. Probabilistically, the above sentence is a logical law because it is *certain* under all assignments of probability functions to its component variables. Certainty, as indicated above, is to be cashed out in terms of having a subjective conditional probability value of 1 for all assignments of probability functions to the component variables. That is:

$$A \text{ is certain means } \forall C [P(A/C) = 1]$$

Applied to the above, the sentence is certain if and only if:

$$\forall C [((P \rightarrow Q) \rightarrow (-Q \rightarrow -P)) / C = 1]$$

The sentence (A) is a logical law of truth-theoretic logic because, for all assignments of truth-values to the sub-sentential components P and Q the truth-value of the overall sentence is always true. That is, the sentence preserves truth under all interpretations of its sub-sentential components. Similarly, with the probabilistic interpretation. The sentence (A) expresses a logical law insofar as it preserves a conditional probability value of 1 for all assignments of probability values to its subsentential components. That is, it preserves a conditional probability value of 1 for all interpretations of its sub-sentential components.

Similar remarks apply to inferences as well as sentences. Suppose we have a sequent expression of the form, $A_1, A_2 \dots A_n \vdash B$. According to Field, such an inference is *legitimate under an interpretation* if and only if: $\forall C [P(B/C) \geq P(A_1 \dots A_n)/C]$. Consider, for example, the corresponding sequent expression of the sentence (A):

$$P \rightarrow Q, -Q \vdash -P \quad (A')$$

On a truth-theoretic interpretation of classical logic, (A') is a logical law because the inference from $P \rightarrow Q$ and $-Q$ to $-P$ necessarily preserves truth. It is impossible to make both premisses true and the conclusion false. On a probabilistic interpretation, the inference is *legitimate under an interpretation* if and only if:

$$(\forall C) [P(-P/C) \geq P(P \rightarrow Q, -Q)/C]$$

And the sentence is probabilistically valid if and only if it is legitimate under all interpretations. That is, legitimate for all assignments of conditional probability values to the sub-sentential components P and Q.

Therefore, on Field's probabilistic interpretation, the notion of logical consequence is defined as follows: A sentence B is a logical consequence of a set of sentences $A_1 \dots A_n$ if and only if the sequent expression $A_1 \dots A_n \vdash B$ preserves a subjective conditional probability value of 1 for all assignments of subjective conditional probability values to the variables $A_1 \dots A_n$.

The space is not available for a detailed exegesis and evaluation of Field's interpretation. Field goes on to derive soundness and completeness proofs, and a recursive definition of the probability of sentences, showing how the conditional probability of complex sentences depends on the conditional probability of their subsentences and subformulas. He also goes on to extend the probabilistic semantics to predicate languages.¹¹ All I have intended to do here is describe the outlines of an approach to providing a formal semantics for logic which more adequately reflects the role of rationality in constitutively governing relations between mental states. Whether one regards Field's approach as ultimately successful or not is independent of my main point which is that a truth-theoretical notion of logic does not

11. "Logic, Meaning, and Conceptual Role", op. cit., pp. 383-88.

adequately capture our intuitive concept of rationality as a principle constitutively governing relations between mental states.

4

If the ER principle is rejected, then rationality can constitutively govern relations between mental states in virtue of their internal as well as external components. That is, relations holding between internal components, or narrow mental states, are also essentially subject to norms of rationality. This has important consequences for the reconciliation of anomalism and supervenience proposed in Part 1 of this Chapter. This reconciliation can be characterized as follows.

- (i) Internal components, or narrow mental states, strongly supervene on physical states in such a way that individuation conditions are transmitted.
- (ii) The physical states which form the realizing bases of internal components are both physical and natural states.
- (iii) Therefore, internal components have individuation conditions which are both natural and physical.

For mental states widely individuated, the situation is correspondingly different:

- (iv) Wide mental states strongly supervene on states in such a way that individuation conditions are transmitted.
- (v) The states which form the realizing bases of wide mental states are not natural physical states. Any specification of their identity conditions which is natural is patently non-physical, and any specification of their identity conditions which is physical is patently non-natural.
- (vi) Therefore, wide mental state do not have individuation conditions which are both natural and physical.

The reconciliation represented in (i)-(vi) depends essentially on the ER principle. Rejection of this principle, then, will have several major consequences.

The most important consequence is that rationality can now also apply to relations between narrow mental states. However, if this rationality is constitutive of relations between internal components, then Davidson's arguments against psychophysical laws can presumably be re-run to establish the anomalism of internal components. Is this compatible with with the claim that narrow mental states strongly supervene on physical, specifically brain, states? By now, the answer to this should be clear: There is an incompatibility here only if the realizing physical or brain states are natural physical properties. That is, only if they have identity conditions which are specifiable in natural physical terms. It is an essential part of the principle of strong supervenience that modal supervenience conditionals transmit individuation conditions. Therefore, if we accept strong supervenience,

the claim expressed in (i) must remain intact. However, it is possible to make the transmission of individuation conditions consistent with the anomalism of narrow mental states by making the following amendment to (ii):

- (iia) The states which form the realizing bases of narrow mental states are not natural physical states.

The states which realize narrow mental states are presumably brain states of some sort; but the notion of a brain state can be understood in two distinct ways. On the one hand it could be understood to mean a well-defined natural property of neurophysiology: that is, a property taxonomized by neurophysiology, and playing an explanatory role in the explanations of neurophysiology - a property which has explanatory relevance of a sort appropriate to neurophysiology. On the other hand, the notion could be understood to mean a state actually instantiated in the brain. And the two notions are far from equivalent.

A state actually instantiated in the brain could be constructed out of natural properties of neurophysiology. And a state which is constructed out of natural neurophysiological properties need not itself be a natural neurophysiological property: it need not figure in the explanations of neurophysiology, and any natural specification of it might not be possible in purely neurophysiological terms. The analogy with the realizing bases of wide mental states is useful here. In this case it was argued that the realizing bases of wide mental states were essentially composite, composed of a brain state b, a state of affairs

f, and a connecting relation R. In connection with this it was argued that the only natural criterion available for unifying b, R, and f into a single property was that they together realize the wide mental state M. That is, the only natural specification of the unitary state [bRf] involved essential reference to the supervenient mental state M. Conversely, any specification which did not involve reference to M - any purely physical specification - was not a natural specification. The suggestion being proposed now is that the situation may be very similar in the case of the realizing base properties of *narrow* mental states. That is, the realizing base property of a narrow mental state might, typically, be composed of such states as $b_1, b_2, b_3 \dots$ where each of these is a natural property of neurophysiology. However, viewed purely from the perspective of neurophysiology, there might be no natural criterion for unifying b_1, b_2, b_3 into the single property [b_1, b_2, b_3]. The only natural rationale underlying such a unification might be that together they realize the narrow mental state M^* . Verification of such a view, of course, would be the work of some future neuroscience. The position is outlined to show what sort of situation is needed if narrow mental states are to be strongly supervenient on physical states without receiving natural physical individuation conditions.

Given (iia), the following amendment to (iii) is also required:

(iiia) Therefore, narrow mental states do not have individuation conditions which are both natural and physical.

I will now go on to consider the consequences principles such as (iia) and (iiia) have for the nature of explanation in folk and cognitive psychology.

Appendix II:

Reply to a Possible Objection: Ontological versus Individuation Dependence.

1

In this appendix the proposed reconciliation of anomalism and supervenience will be defended against a possible objection.

The basis of the objection is that the reconciliation leaves us with an inadequate account of the supervenience relation. The relation implicated in the proposed reconciliation is the supervenience relation by name only; it is not a genuine relation of determination. The primary aim of the supervenience claim, in the mind-body case, is to account for the *dependency* of mental properties on physical properties. But, so the objection runs, if we gerrymander the supervenience base, reading it, as it were, downwards from the mental as the mental occurs in any possible world, it is not at all clear that we get such dependence. The relation of dependence, if anything, seems to run in the opposite direction.¹

The basic problem with this objection, it will be argued, is that it confuses what I shall call *ontological* dependence with *individuation*

1. This objection was pointed out to me by Simon Blackburn in his capacity as editor of *Mind*.

dependence. Ontological dependence is to be understood as a brutemetaphysical relation of dependence, such as the supervenience relation, or, perhaps, the causal relation, purports to be. In contrast, I will say that a property P is individuation dependent on a property Q if and only if reference to Q is an essential part of the specification of the identity conditions of P. Clearly, more could be done to sharpen up these characterizations, but this level of precision will suffice for present purposes.

According to the proposed reconciliation of the principles of anomalism and supervenience, supervenience (and realizing) base properties are individuated in terms of the supervenient properties they underlie. Thus the relation of individuation dependence is clearly top-down. Nevertheless, it does not follow that the relation of ontological dependence is similarly top-down. The following remarks will be confined to realizing base properties, but they apply equally to supervenience base properties.

To make the above remarks clearer, we can imagine two sets of properties, set A = {A₁ ... A_n}, and set B = {B₁ ... B_n}. What is important here is that the properties of A, and those of B, are each individuated according to standards, rules, and principles that are peculiar to A and B respectively. The individuating standards, rules, and principles of set A are distinct from those of set B. The specific nature of these standards, rules, and principles is unimportant here, though it has been claimed that they will centre around the nature of the explanations of A and B respectively.

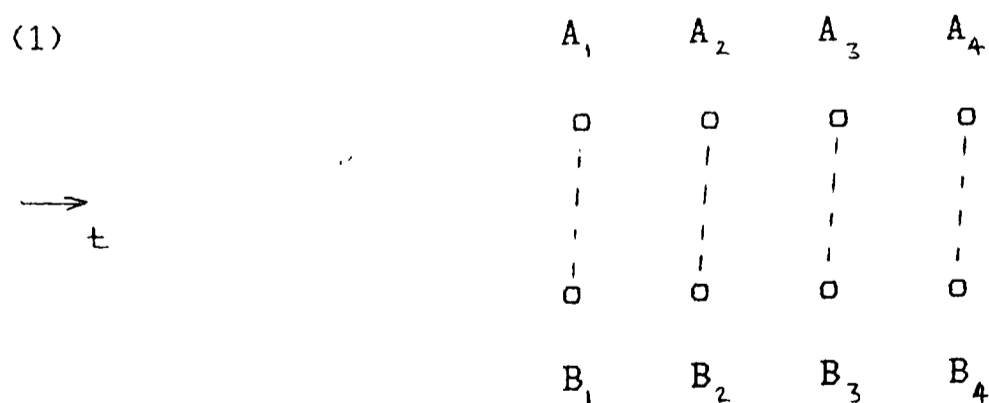
Within each set, the individual properties stand in certain relations to one another, and in doing so, form a *system* of properties. For the sake of representational simplicity, assume the only relations involved here are ones of linear succession through time. Then it would be possible to represent a system, for example the system formed from set A, as follows:



where $A_1 \dots A_4$ are properties of A, t is the time axis.

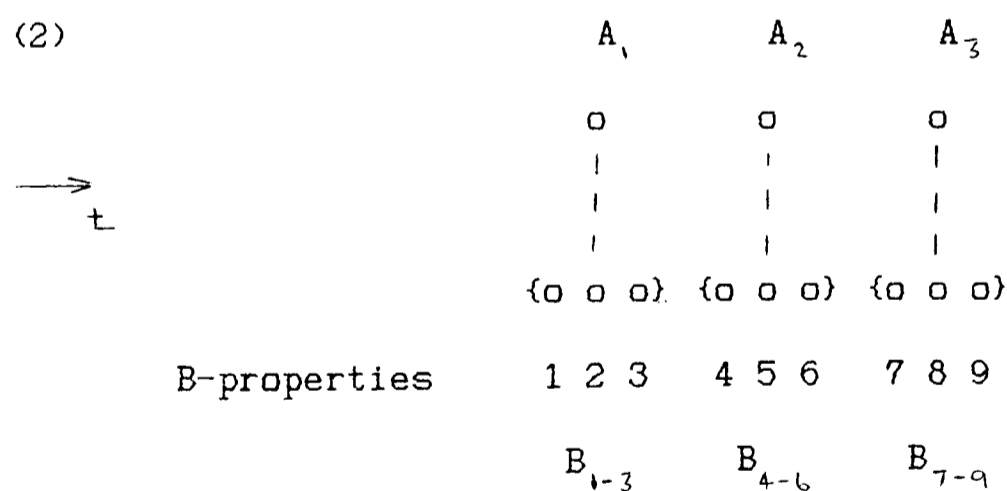
Assume, now, that A-properties stand to B-properties in the realization relation, that is, that B-properties are realizing bases of A-properties. Given this relation, it is possible to imagine, it is possible to imagine three types of variation on this relation.

In the first type of case, the properties in each of the two systems formed from A and B can be related by way of a one:one mapping function. That is:



In this type of case, no relation of individuation dependence need hold between A-properties and B-properties. Each property in A will be individuated according to standards, rules and principles appropriate to A. And each property in B will be individuated according to standards, rules and principles appropriate to B. Thus, no property in A is individuation dependent on any property in B, and vice versa. However, this is compatible with their being ontological dependence of A-properties on B-properties. It might be possible, for example, to say such things as: 'if any object/organism were to possess B₁, then it must also possess A₁.' Or, perhaps, 'an object cannot change in point of A-properties without also changing in point of B-properties.' Ontological dependence is distinct from, and independent of, individuation dependence.

The second type of case consists in a more complex one: many mapping from A-properties onto B-properties. That is:



where B₁₋₃ is the property formed by the conjunction of properties B₁-B₃; similarly for B₄₋₆ and B₇₋₉.

The relation of individuation dependence is here more difficult to discern. On the one hand, it might be that the rationale for conjoining properties $B_1 \dots B_n$ into the single property B_{1-n} can be found within the individuating standards, rules and principles of the set B itself. In this case, the property B_{1-n} would be a natural property of B ; it would not be gerrymandered. If this were so, then the situation represented in (2) would be relevantly similar to that represented in (1), insofar as there need be no relation of independence running from A -properties to B -properties (or, indeed, in the other direction).

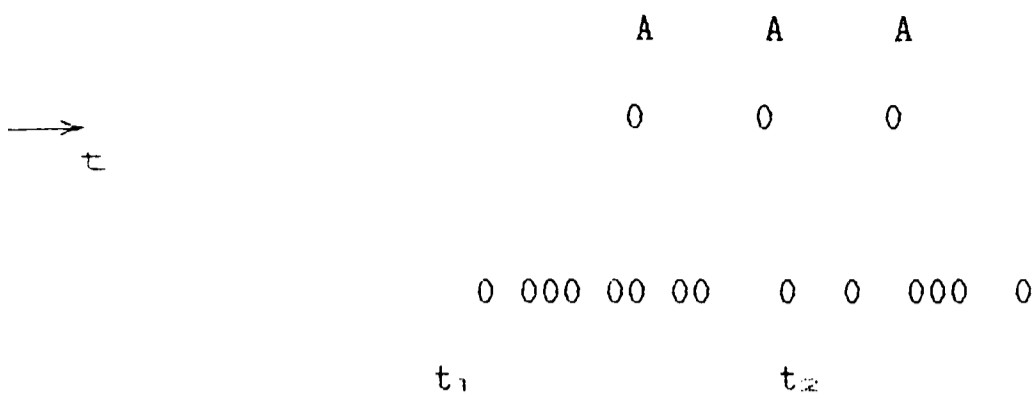
However, there is another possibility. The rationale for conjoining the properties $B_1 \dots B_n$ into the single property B_{1-n} might *not* be found in the individuating standards, rules and principles of the set B . That is, there might be nothing in those standards, rules and principles which could sanction, legitimize, or explain such a conjunction. In this case, B_{1-n} would not be a natural property of B , and the rationale for conjoining $B_1 \dots B_n$ would have to be sought outside B . It is clear where such a rationale would derive from. $B_1 \dots B_n$ would be unified into B_{1-n} because B_1 , b_2 , and B_n , together, realize A_1 . Thus, the relation of individuation dependence is clearly top-down. Specification of the identity conditions of B_{1-n} , if this is not to appear a hopelessly gerrymandered property, must make essential reference to A_1 .

However, such top-down individuation dependence is perfectly compatible with bottom-up ontological dependence. For, even if the complex property B_{1-n} is individuated in terms of A_1 , it is still true

that A_1 is realized by B_1 , B_2 and B_3 entering into appropriate relations with one another at the B-level. It is still true, for example, that if any object were to possess B_1 , B_2 and B_3 entering into those relations circumscribed by the conjunction of those properties, then it must also possess A_1 . This is a different issue altogether from the question of how the complex property B_{1-3} is to be individuated. Top-down individuation dependence of B_{1-3} on A_1 is compatible with bottom-up ontological dependence of A_1 on B_1 , B_2 , and B_3 .

In the situation described in (2), the components of the realizing base B_{1-3} are properties which are identifiable according to the individuating standards, rules and principles of set B. However, it is possible to imagine a situation in which even this condition is not met. The envisaged situation is one in which the two systems formed from A and B are radically disjoint with respect to one another.² There is no possible mapping - either one:one or one:many of the properties of A onto the properties of B. Davidson's thesis of the anomalism of the mental might plausibly be construed as asserting such essential disjointness between mental and physical properties. The situation here could be represented as follows:

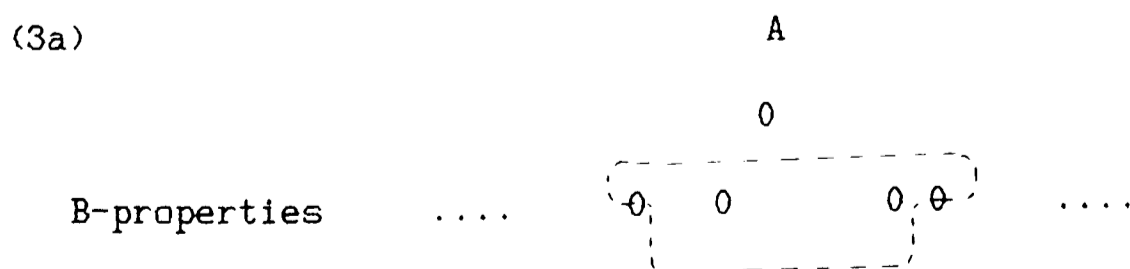
2. I take it that the idea of such a radically disjoint system of properties is now fairly familiar. One of the originators of the idea was Fodor, "Special Sciences", op. cit. The idea is also discussed in D. Leder, "Troubles with Token Identity", in Philosophical Studies, vol.47, no.1 (1985), pp. 79-94. Also Jennifer Hornsby, "Physicalism, Events, and Part-Whole Relations", in LePore and McLaughlin eds., Actions and Events: Perspectives on the Philosophy of Donald Davidson, op. cit., pp. 444-58. Both Leder and Hornsby argue, correctly in my view, that such disjointness of properties precludes the possibility of those properties sharing instances, i.e. would rule out the possibility of token identity holding between instances of those properties.



That is, (3) is meant to represent a situation a situation in which there is no mapping of properties of A onto, as it were, 'whole numbers' of B-properties.

The first point to note is that such disjointness is perfectly compatible with the determination of A-properties by B-properties. For example, one might take a *time-slice* out of the linear succession of the properties represented above, say between times t_1 and t_2 . Then it would be possible to say that any two objects which are exactly alike with respect to the B-properties occurring in the time-slice t_1 - t_2 must also be exactly alike with respect to the A-properties occurring between t_1 and t_2 . That is, the mutual disjointedness of the systems formed out of A and B does not rule out the *global* supervenience of the system of properties formed out of A on the system formed out of B.

However, the case need not rest on considerations of global supervenience. The picture represented in (3) can be narrowed down to a *local* level, perhaps as follows:



That is, we might find that the taxonomic apparatus of the system A so radically cross-cuts that of the system B that the portion of B which realizes A_1 does not even correspond to a 'whole number' of those properties individuated by the taxonomic apparatus of B. In this case there would be no temptation to regard the portion of B enclosed by the dotted line as a natural property of B. The rationale for isolating that portion of B, the only rationale for *drawing the dotted lines* is to identify that portion of B which is responsible for A_1 . Thus the relation of individuation dependence is clearly top-down. Nevertheless, just as with the situation represented in (2), the relation of ontological dependence is still clearly bottom-up.

To see this, imagine two objects possessing through time the sequence of B-properties occurring between times t_1 and t_2 . Given this, it would then be possible to imagine a period of time when both objects possessed that portion of B which is enclosed within the dotted lines of (3a). Then, it would be possible to say that these two objects must also possess A_1 . There still exists an ontological dependency of A_1 on that enclosed portion of B, even if such a dependency cannot be expressed in the vocabulary of B. Ontological dependence can be bottom-up even if individuation dependence is top-down.

An example might make this clearer. Consider the earlier example of moral supervenience. In this example, it was argued that a certain moral property - the property of being good - could be realized by what was an obviously gerrymandered physical property. The specific property chosen was, in this case, formed from the conjunction of the property of

being certain movements of the arms, legs, and torso, towards a specified object (a child), in an appropriate medium (water). The realizing base is obviously not a natural property of physical theory. Hence, any natural specification of the property involved essential reference to the supervenient moral property. Thus the relation of individuation dependence is top-down. But this does not alter the fact that the moral property of being good is dependent on this conglomeration of physical properties. If any two objects were to instantiate the above physical properties, then they must also instantiate the moral property. The dependence of the moral property on the physical properties does not alter, even if the the gerrymandered construction of physical properties which actually realizes the moral property can only be specified in terms of that moral property. Top-down individuation dependence does not undermine bottom-up ontological dependence.

What is crucial in demonstrating the distinctness of ontological and individuation dependence is the distinction between natural and gerrymandered properties of a domain of explanation. Consider again the two systems formed out of the sets A and B. The salient point is this: just because it is a gerrymandered property of B which realizes any given property of A, this does not mean that the properties of B are all individuation dependent on properties of A. As was stated, the two sets of properties are individuated according to their own, distinct, taxonomic apparatus. This allows A-properties to be ontologically dependent on the natural properties of B, while at the same time,

allowing that the gerrymandered properties of B to be individuation dependent on the A-properties they realize.

Appendix III:

The Nature of Psychological Explanation: Normativism Versus Generalism

1

The explanatorily narrow interpretation of the dual component theory identifies internal components or narrow mental states as loci of causal/explanatory relevance. This can lead to two possible views. According to one, ordinary wide mental states possess causal/explanatory relevance, but only in virtue of being partly composed of narrow mental states which are the primary source. According to the other, both internal components and ordinary mental states possess causal/explanatory relevance, where the latter need not be entirely dependent on the former, but where only the former possesses relevance of a sort appropriate to cognitive psychology. For present purposes, this distinction is not important. Both types of case lead to the possibility of an application of Ockham's razor and the consequent factoring off of on the one hand, (i) the explanations of a folk psychology dealing with relations between ordinary wide mental states, and (ii) the explanations of a scientific psychology, dealing with interactions between *narrow* mental states.

The purpose of this appendix is to develop the reconciliation of anomalism and supervenience described in the previous Chapter from the

point of view of its implications for the nature of explanation in each of (i) and (ii). The focus will be on the issue of whether the explanations in each of (i) and (ii) must be viewed as generalist or normativist explanations. A useful point to begin will be by returning to Davidson's arguments against psychophysical laws, and, specifically, contrasting this view with that of *functionalism*. My exegetical vehicles here will be Brian Loar's Mind and Meaning, and John McDowell's reply given in "Functionalism and Anomalous Monism".¹

Davidson's argument against psychophysical laws hinges on the claim that the patterns implicated by the norms of rationality cannot be expressed in non-intentional terms. It is Loar's contention on the other hand, that these patterns can be given a functional interpretation, where the type of functionalism at issue here involves topic-neutral and, hence, non-intentional specification of mental states.² Loar's attempted functional interpretation of rationality centres around two sorts of constraint. Firstly, *L-constraints*, which cover relations between mental states which hold in virtue of, and hence are specifiable in terms of, their logical forms.³ Secondly, *M-constraints*, which cover relations between mental states which hold in virtue of, and hence are specifiable in terms of, the content of those mental states.⁴ It is the L-constraints which represent *deductive*

-
1. Brian Loar, Mind and Meaning, (Cambridge, Cambridge University Press, 1981). John McDowell, "Functionalism and Anomalous Monism", in LePore and McLaughlin eds., Actions and Events: Perspectives on the Philosophy of Donald Davidson, op. cit., pp. 387-98.
 2. Mind and Meaning, op. cit., p. 21.
 3. Mind and Meaning, op. cit., pp. 71-74.
 4. Mind and Meaning, op. cit., pp. 81-85.

rationality, and following McDowell, discussion will be largely limited to the L-constraints. Results applicable here should also be applicable to M-constraints.

According to McDowell, deductive rationality can be regarded as;

" ... a capacity, more or less perfectly instantiated in different rational individuals, to hold beliefs when, and because, they follow deductively from other beliefs that one holds. If we are careful, it need do no harm to picture a particular instantiation of deductive rationality as a more or less approximate grasp of a normative structure, determining what follows from what and thus what ought to be believed, given other beliefs, for deductively connected reasons. The Davidsonian claim, now, is that this structure (if we allow ourselves that picture) cannot be abstracted away from relations between contents, or forms of content, in such a way that we might hope to find the abstracted structure exemplified in the interrelations among a system of items described in non-intentional terms."⁵

One must be careful here. McDowell's comments concerning 'abstracted structure' refer to the structure of (deductive) reason itself. And the normative character of reason here is crucial. The McDowell/Davidson claim is *not* that relations between mental states cannot be "exemplified in the interrelations among a system of items described in non-intentional terms". Rather, it is the structure of reason which cannot be so exemplified; it is what makes the relations between mental states instances of rational relations which cannot be thus exemplified. The reason is that if the relations between mental states were interpreted in these non-intentional terms, then the *normative* character of rationality would be lost. Since it is as a normative notion that

5. "Functionalism and Anomalous Monism", op. cit., p.388.

rationality has constitutive status, any non-intentional interpretation of rationality would rob it of its constitutive status.

Though McDowell does not talk in these terms, crucial to understanding his argument against Loar is the distinction between *normative* and *descriptive* notions of rationality. Consider how one might try to characterize the notion of rationality. We might try describing the actual psychological profile of an organism which we regard as rational, noting the transitions which this profile undergoes through time etc. Then, abstracting across different descriptions of different, supposedly rational, organisms, we arrive at a description of some "... minimally necessary structure exemplified in the actual psychological economy of anything that could be recognized as a rational mind."⁶ However, any temptation to regard this described structure as the structure of rationality must be based on the view that rationality is a descriptive, rather than a normative, concept. When Davidson argues that the structures required by rationality cannot be given a non-intentional interpretation, the notion of rationality at work is a normative notion. Rationality is viewed as an ideal which specifies what states an organism *should* be in conditionally on being in other specified states. This ideal may never actually be exhibited in the psychological profile of any organism.

6. "Functionalism and Anomalous Monism", op. cit., p. 389.

Loar's claim, then, if it is to pose any threat to Davidson, must be that a functional interpretation of rationality is possible, even when rationality is construed as a normative principle. McDowell levels two charges against this claim.

The first, and of lesser importance, concerns the *scope* of the constraints provided by Loar. These cover "... at best a fragment of rationality in general".⁷ Indeed, Loar himself points out that these constraints are quite undemanding. For example, they do not ensure even a rudimentary proficiency at making, or assessing, inferences. Therefore, there is no way that the L-constraints can be thought of as capturing the structure of deductive reason itself - for this structure is normative, and must reflect what in general follows from what.

Secondly, and more importantly, McDowell argues that a functional interpretation of the sort proposed by Loar could never be anything more than a descriptive account of "... certain transitions and refrainings from transitions that minds are as a matter of fact prone to."⁸ That is, the functional interpretation could never be anything more than a descriptive account of the actual psychological profile of various organisms. The functional account gives us no understanding of notions such as deductive consequence, rational cogency, etc. The functional interpretation can, at most, *describe* various transitions which minds

7. "Functionalism and Anomalous Monism", op. cit., p. 387.

8. "Functionalism and Anomalous Monism", op. cit., p. 390. Substantially similar arguments can be found in Kathleen Lennon, "Anti-Reductionist Materialism", *Inquiry*, vol.27, pp. 363-80.

that we regard as rational tend to undergo. It can in no way elicit any principles underlying these transitions.

Therefore, any functional (non-intentional) interpretation of the principle of rationality could not capture the normative character of rationality. However, according to Davidson, it is as a normative notion that rationality possesses constitutive status. Therefore, if Davidson is correct, any non-intentional interpretation of rationality must fail to capture the constitutive status of rationality.

2

If McDowell is correct, then Loar's claim to have given a functional interpretation of the principle of rationality rests on a confusion about the notion of rationality which is deemed to have constitutive status. If the notion of rationality at work was a descriptive one, then a functional interpretation of rationality constraints would be possible. However, the appropriate notion of rationality is, in fact, normative. According to McDowell, Loar's confusion on this matter is symptomatic of a "... prejudice about the character of understanding we can achieve by employing the conceptual apparatus that is governed by the constitutive force of rationality."⁹

9. "Functionalism and Anomalous Monism", op. cit., p. 389.

The relevant distinction, here, is between "... explanations in which things are made intelligible by being revealed to be, or to approximate to being, as they rationally ought to be...", and explanations in which "... one makes things intelligible by representing their coming into being as a particular instance of how things generally tend to happen...".¹⁰ Using a terminology which is not McDowell's, I will call explanations of the former sort *normativist* explanations. Those of the latter sort I will call *generalist* explanations.

A generalist explanation makes an event intelligible by identifying it as a particular instance of how the world generally works.¹¹ The underlying assumption here is that the world develops according to an order which is causal and, above all, regular (even if the regularity here is only probabilistic). An event is made intelligible by a generalist explanation when it is shown to be part of the unfolding of this regular order.

If Davidson/McDowell are correct, ordinary belief-desire explanation cannot be *correctly* viewed as generalist explanation. Nevertheless, such explanations can *consistently* be viewed as generalist in character, and many have attempted to view them in this way. I will now briefly consider how belief-desire explanations can be fitted into

10. "Functionalism and Anomalous Monism", op. cit., pp. 389-90.

11. c.f. Philip Pettit, "Broad-Minded Explanation and Psychology", in P. Pettit and J. McDowell eds., Subject, Thought and Context (Oxford, Oxford University Press, 1986), pp. 17-58, p. 25.

the generalist mould. While this model will eventually be rejected for belief-desire explanations, it will serve to bring out an important point concerning generalist explanation and causal/explanatory relevance.

The model of belief-desire explanation provided by the generalist framework is as follows. Firstly, folk psychology describes various principles linking beliefs, desires, hopes, fears, expectations etc. On the generalist account, these principles will be understood as at least providing evidence for, though not necessarily to be expressions of, determinate and regular patterns in nature. When one of these principles is applied in the explanation of an action, the instantiation of the antecedent explains the instantiation of the consequent in virtue of showing that this instantiation is part of an ordered regular pattern which occurs in nature. Moreover, the instantiation of the antecedent, a conglomeration of mental states, is taken to be the cause of the action explained. It has causal relevance.¹²

However, there are well known problems with regarding action explanation as causal explanation, and, hence, over regarding intentional profiles as bearers of causal relevance. These problems concern the independent verifiability of antecedent and consequent in this type of explanation.¹³ To appreciate this it is important to realize the *a priori* character of the principles involved in action

12. My account here owes much to the clear exposition of Pettit, *op. cit.*

13. This point is made by Pettit, *op. cit.*, pp. 27-28. See also C. Peacocke, Holistic Explanation (Oxford, Oxford University Press, 1981), pp. 147-55.

explanation. If someone desires that p, believes that not p, and believes that by A-ing he can bring it about that p, then, *ceteris paribus* he As. Principles such as this are constitutive of our conception of what it is to believe and desire things; understanding that conception is sufficient for seeing that the principle is true. Hence, such principles are not established empirically; they are not *parts of* our ordinary folk psychological explanation schema, rather they are *conditions* of it. As such, they are a priori.¹⁴

The a priori character of the principles creates a problem: failure of the consequent to follow cannot be allowed to falsify the principle, it must be taken to show that the antecedent did not, in fact, obtain. That is, the antecedent of any principle underlying belief-desire explanation is not verifiable independently of the consequent.¹⁵ This lack of independent verifiability creates a problem for those who wish to regard action explanation as causal explanation, and hence, mental properties as the bearers of causal relevance.

The resolution of this problem has been sought by many in cases thought to be analogous to action explanation. These supposedly analogous cases concern dispositional properties. Consider, for example, the property of being fragile.¹⁶ There is a conceptual connection between fragility and breaking which makes explanations of

14. "Broad-Minded Explanation and Psychology", op. cit., pp. 26-27.

15. Peacocke, Holistic Explanation, op. cit., pp. 147-55. See also Pettit, "Broad-Minded Explanation and Psychology", op. cit., pp. 27-29.

16. The example is taken from Pettit, "Broad-Minded Explanation and Psychology", op. cit., pp. 27-29.

the latter in terms of the former suspiciously convenient. Contrast this with an explanation of the breaking in terms of molecular structure. Molecular structure can be determined without reference to whether the object breaks under severe shocks, and there is nothing suspicious about the regularity which the principle invoking it reveals in the breakage.

The resolution now becomes clear. The explanation by means of the a priori principle is not to be regarded as an ideal generalist account, but, rather, one which is essentially constrained by considerations of ignorance. An ideal generalist explanation of an object breaking would presumably be cast at the level of molecular structure. Referring to fragility is a way of referring to whatever intrinsic property it is - say, having such and such a molecular structure - that causes things to shatter under certain conditions of stress.¹⁷

A similar line can be taken by someone who wishes to regard action explanation as generalist explanation. According to this view, ordinary folk psychological explanations also are motivated by considerations of ignorance. In particular, the explanatory force of any folk psychological state derives from the fact that it goes proxy for some intrinsic property of the agent. This intrinsic property will be some neurophysiological state or complex, and the ideal generalist account will be cast at this level. And the explanatory force of folk

17. "Broad-Minded Explanation and Psychology", op. cit., p. 29.

psychological explanations derives from the fact that the states they invoke go proxy for underlying neurophysiological states, and stand to these states in the same relation that fragility stands to molecular structure.¹⁸ Hence, it is in virtue of this relation that mental properties possess causal relevance.

There are two essential features of this solution, one concerning the relation between mental properties and the underlying genuinely explanatory states, the other concerning the nature of those underlying states.

Firstly, the relation. This is clearly one of *supervenience* or *realization*. The dispositional property of being fragile is realized by a certain molecular structure. The relation here is one of conditional dependence. The biconditionality of the relation is ruled out by the possibility of variable realization. Similarly, the intentional profile, which forms the antecedent of a given belief-desire principle, is realized by an underlying neurophysiological state or complex. Again, variable realization blocks biconditional correlation. Therefore, according to the generalist model as applied to mental states, supervenience is an essential part of the account of how mental states possess causal relevance.

Secondly, the nature of the underlying state. Consider the case of fragility. On the account being considered, the property of being

18. "Broad-Minded Explanation and Psychology", *op. cit.*, p. 29.

fragile possesses causal relevance - its ability to play an explanatory role in causal explanations - because of its realization by some underlying state, a molecular structure for example. Now, the essential feature of generalist explanations is that they function by exhibiting regularity. Therefore, if the explanatory value of fragility is to ride on the explanatory value of an underlying molecular state, then that underlying state must be such that it is capable of exhibiting regularity. That is, it must figure in generalist explanations. Therefore, assuming those explanations are causal, the underlying molecular state must possess causal relevance. However, if a molecular structure or property possesses causal relevance, then that property is a *natural* property relative to the domain of explanation concerned with molecular explanation. This follows from the earlier characterization of naturalness.

Therefore, applying the same model, if mental states are to possess causal relevance within the generalist framework, they must *supervene* on (presumably physical) properties which are *natural* (relative to the domain of neurophysiological explanation).

3

Generalist explanation makes an event intelligible by representing it as part of a regular succession instantiated in the world. However, not all explanations function in this way. Some explain an event in relation to a *norm* at which the world, or some part of the world, aims.

In such a case, the event would be made intelligible by being represented as something that had to happen if the world was to continue to satisfy that norm. Such an explanation would be *normativist*.

Normativist explanations are to be found only in domains where selection, design, or some similar constraint, has determined that the systems there satisfy this or that norm. In particular, according to Pettit, three conditions must hold. Firstly, there must be an adequately demarcated domain of explanatory concern. Secondly, the systems in that domain are designed or selected so as to fit, more or less precisely, a certain conception. Thirdly, we understand that conception well enough to be able to spell out the norms which constitute that conception.¹⁹

Pettit argues that these conditions are fulfilled in the action case, and, consequently, that action explanation in terms of beliefs, desires... etc. can consistently be depicted as normativist in character. Pettit goes on to argue that such explanation *must* be regarded as normativist. The reason for this stems from the fact, expressed earlier, that the principles engaged in, and underlying, action explanation are a priori in character. While this is not logically inconsistent with a generalist account, it does lend more support to a normativist account. For, if a generalist model is adopted, the a priori status of these principles goes unexplained. In

19. Pettit, "Broad-Minded Explanation and Psychology", op. cit., p. 41.

contrast, it becomes readily intelligible on a normativist view of action explanation. On the normativist account, but not on that of the generalist, the principles of action explanation relating beliefs, desires etc. to behaviour are *norms* which express our conception of what it is to be a human agent. These principles are essential preconditions of any normativist, but not of any generalist, explanation. They are constitutive of the domain of explanatory concern, hence must be a priori.²⁰

4

The considerations expressed earlier, those concerning the derivation of causal relevance on a generalist model lend support to Pettit's claim. If a generalist model of action explanation is adopted, it was argued that two conditions need to be satisfied in order for action explanation to be regarded as genuine causal explanation, hence, for mental state to be regarded as genuine bearers of causal relevance:

- (i) The mental states which form the antecedent of a particular explanation must supervene upon, or be realized by, underlying physical properties.
- (ii) These underlying physical properties must be *natural*.

20. Pettit, "Broad-Minded Explanation and Psychology", op. cit., p. 45.

Consider, firstly, the case of *wide* mental states and their realizing bases. In Chapter IV, it was argued that, with regard to the realizing bases of wide mental states, the conditions expressed in (i) and (ii) could not be simultaneously fulfilled. The realizing base property of a wide mental state could not plausibly be viewed as both physical and natural. Any physical specification is patently non-natural, and any natural specification is patently non-physical. Therefore, either one rejects the claim that wide mental states have causal/explanatory relevance of any sort, or one must reject the generalist account of belief-desire explanation. Given an available alternative in the normativist account, the more favourable option is surely the latter.

The situation is more complicated in the case of *narrow* mental states. The important point is, however, that the principles elicited in chapters III and IV, regarding the reconciliation of anomalism and supervenience, entail that any explanation involving narrow mental states cannot be generalist in character.

If rationality applied to mental states purely in virtue of their external components, then it would be possible for narrow mental states to be realized by natural physical properties. And this would render possible what is, in many ways, an attractive picture of a scientific cognitive psychology. For, if this were the case, it would be possible to maintain that, although the explanations of folk psychology are irredeemably normativist, the explanations of a respectable cognitive science, dealing with relations between narrow mental states, are

generalist (and, thus, in some sense, 'properly scientific'). Perhaps this respectable cognitive science would be in many important ways a *functionalist* account; for given that rationality does not constitutively govern relations between narrow mental states, it would be possible for these states to be given a functional interpretation. Most importantly, perhaps, this respectable cognitive science would have no normative constitutive principle of rationality to deal with. It could, therefore, have as its goal explication of all those 'minimally necessary structures' which are "... exemplified in the actual psychological economy of anything that could be recognized as a rational mind."²¹

However, if the arguments given in chapters III and IV are correct, then this picture of the relation between folk and cognitive psychology cannot be sustained. For, it has been argued that rationality cannot plausibly be viewed as applying to mental states purely in virtue of their external component. Therefore, narrow mental states cannot supervene on states which are both physical and natural. And if this is so, the possibility of a generalist explanatory model being applied to narrow mental states suffers from exactly the same problems as when applied to wide mental states: A necessary condition of explanation involving narrow mental states being generalist explanation is that these states supervene on natural physical properties. But given the correctness of the arguments of chapters III and IV, such a condition

21. McDowell, "Functionalism and Anomalous Monism", op. cit., p. 389.

cannot obtain. Therefore, narrow mental states do not figure in generalist explanations.

For this reason, the picture of the 'respectable scientific psychology' described above must be rejected. The picture presupposes an untenable account of the relation between the principles of anomalism and supervenience as they apply to narrow mental states.

Chapter V:

Formal Explanatory Methodological Solipsism: The Role of Syntax in Psychological Explanation

1

It was argued in Chapter I, that the principle of Explanatory Methodological Solipsism (EMS) could be subdivided into the principles of Formal and Conceptual Explanatory Methodological Solipsism (FEMS and CEMS, respectively) depending on one's view of which properties of narrow mental states have causal/explanatory relevance. There are two possible answers to this question: (a) narrow semantic properties, and (b) syntactic properties.

To be more precise: In this thesis, the internal (and the external) component of a mental property has been regarded as an $(n+1)$ -order property possessed by an n -order mental property. Then, the narrow semantic properties, and the syntactic properties, are to be regarded as $(n+2)$ -order properties attaching to an $(n+1)$ -order internal component, and, hence, to an n -order mental property. Narrow semantic properties, and syntactic properties, attach primarily to narrow mental states, and to ordinary wide mental states derivatively upon this. Thus both types of property constitute ways of picking out the internal component.

The claim that it is the narrow semantic properties of mental states which possess causal relevance, and hence figure in the

explanations of psychology, was discussed in Chapter II. There it was argued that this claim, a statement of CEMS, must be rejected on the grounds that the notion of narrow semantic content is incoherent, and hence that there can be no such things as *narrow content sentences* capable of forming the relata of psychological explanations. More generally, it was argued that there could be no identification of narrow semantic properties independently of the ordinary wide semantic properties of which they form a proper part. While it may be true that narrow semantic properties are a necessary (constitutive) part of (wide) semantic properties, they cannot be abstracted from those semantic properties in such a way as to be capable of figuring in the explanations of psychology.

The other claim to be entertained is that it is the syntactic properties of mental states which are the bearers of causal/explanatory relevance of a sort appropriate to cognitive psychology. This claim is a statement of FEMS. It will be argued that this claim must be rejected, and for similar reasons. That is, it will be argued that there can be no identification of syntactic properties, of any sort useful to psychology, independently of ordinary wide semantic properties. This rejection, if correct, is a result of considerable importance. For if neither narrow semantic nor syntactic properties have causal relevance of a sort appropriate to psychology, then both CEMS and FEMS must be rejected. And this adds up to a rejection of the principle of Explanatory Methodological Solipsism *tout court*.

The claim that it is the syntactic properties of mental states which have causal (and hence explanatory) relevance is a statement of the principle of formal explanatory methodological solipsism (FEMS). It also entails what has become known as the *Syntactic Theory of Mind* (STM). The syntactic theory of mind is a psychological meta-theory; a theory which attempts to delineate the form that theories of cognitive psychology should (or must) take. The best way to understand the import and implications of the syntactic theory is to compare it with two other competing psychological meta-theories. These are, the *Strong Representational Theory of Mind* (Strong RTM) and the *Weak Representational Theory of Mind* (Weak RTM).¹

These various psychological meta-theories can be characterized in terms of the relation they envisage holding between the causally relevant properties of a mental state and the semantic properties of that mental state. Consider, to begin with, the following remarks of Fodor:

" The paradigm situation — the grist for the cognitivist's mill — is the one where propositional attitudes interact causally and do so in virtue of their content If there are true contingent counterfactuals which relate mental state tokens in virtue of their contents, that is presumeably because there are true, contingent

1. The classificatory system adopted here is due to Stich: From Folk Psychology to Cognitive Science, op. cit., chs. 7-9. I will go on to argue that Stich's classification is deficient in several respects.

generalizations which relate mental state types in virtue of their contents."²

Passages such as this describe Strong RTM. What is distinctive of Strong RTM is that it posits a literal identity between semantic properties and the causally relevant properties of (cognitive) psychology: It is the semantic properties associated with a given mental state which are the causally relevant properties of that state. Any cognitive psychology which cleaved to Strong RTM, then, would have to maintain that the laws, generalizations, and explanations of cognitive science must advert to the semantic content of mental states. In other words, cognitive science, in the Strong RTM mould, seeks laws, generalizations, and explanations which relate semantic contents, or sentences expressing semantic contents, and which relate mental properties derivatively upon relating these contents.³

In the above passage, Fodor shows no awareness of the distinction between causal power and causal relevance. His conflation of these two notions is marked, as is often the case, by an insufficiently discriminating use of the expression 'in virtue of'. Thus, when Fodor talks of propositional attitudes interacting causally in virtue of their contents, one interpretation of this claim might run as follows: Propositional attitude tokens have causal *efficacy* (and hence enter into

2. "Propositional Attitudes", The Monist, vol.61, no.4, pp. 501-23. Reprinted in Representations, op. cit., pp. 225-53.

3. This is a modified version of Stich's characterization. The modification consists in elimination of the troublesome phrase 'in virtue of'. See, From Folk Psychology to Cognitive Science, op. cit., pp. 129-30.

token causal relations). They have causal efficacy because they are instances of types characterized by a certain content. And this content has causal *power* 'in virtue of which' it bestows on the propositional attitude tokens the property of causal efficacy. However, this is one interpretation of Fodor's claim which I do not endorse. And there is nothing in my interpretation of Strong RTM which commits it to the implausible claim that contents, or content-sentences, have causal power. The generalizations and explanations in which these contents, or content-sentences, figure may be causal. But this indicates only that they have causal relevance, not causal power. I will eventually defend a version of Strong RTM.

As Stich points out, Fodor's adoption of Strong RTM is not entirely consistent; his position tends to oscillate between Strong RTM and what Stich calls Weak RTM.⁴ Consequently, another passage from Fodor can be used as a characterization of Weak RTM.

" This is the theory:

- (a) Propositional attitude states are relational.
- (b) Among the relata are mental representations...
- (c) Mental representations are symbols; they have both formal and semantic properties
- (d) Mental representations *have their causal roles in virtue of their formal properties.*"⁵

4. This is pointed out by Stich, From Folk Psychology to Cognitive Science, op. cit., 187-91.

5. Representations, op. cit., p.26. Emphasis is Fodor's.

Again, in claim (d) there is a fairly serious conflation of causal power and causal relevance. This will assume importance later.

Weak RTM keeps the claim of Strong RTM that mental states are relations between organisms and contentful or semantically interpreted mental representations [claims (a)-(c) above]. However, it drops Strong RTM's insistence that the generalizations which describe the causal interactions among mental states apply to those states in virtue of their contents. Instead, Weak RTM holds that the generalizations and explanations of cognitive science will be purely formal, describing relations holding between the formal or syntactic properties of mental state types.

Again, I will interpret this as a claim about causal relevance rather than causal power. The claim of Weak RTM, then, is that it is the formal or syntactic properties of mental states which have causal relevance. Thus, although mental states have semantic properties, Weak RTM drops the claim of identity of semantic properties with causally relevant properties. According to Weak RTM, it is syntactic properties which are causally relevant.

There are two importantly distinct interpretations of Weak RTM possible.⁶ Weak RTM claims that the syntactic properties which will be subsumed under the generalizations and explanations of psychology

6. This is pointed out by Stich. From Folk Psychology to Cognitive Science, op. cit., p. 186.

must have associated with them content or semantic properties. However, this last doctrine can be unpacked in two different ways, one of which is considerably stronger than the other. The weaker version claims only that every token mental state to which a cognitive theory applies has some content or other. Such a view would still be a powerful constraint on psychological theorizing, since it rules out cognitive theories which apply to mental states to which no content can be ascribed.⁷ The stronger version of the doctrine makes the additional claim that the semantic properties possessed by a mental state are correlated with the syntactic properties of that state. This is not to be interpreted as a claim of biconditional correlation - a given semantic property can have diverse syntactic realizations - but, rather, one of supervenience. Thus, on the stronger reading of the doctrine, if a pair of mental state tokens are of the same syntactic type, then they must be of the same semantic type also. Since according to Weak RTM, it is the syntactic properties of mental tokens that determine their interactions with other tokens, the view that content is correlated with syntactic type entails that mental states which behave in the same way (i.e. mental states which are functionally identical) must have the same content. Advocates of Weak RTM generally presuppose the supervenience thesis, and I will take this to be an integral part of Weak RTM. Of course, the Twin Earth cases clearly show that ordinary (i.e. wide) semantic content does not supervene on function. Therefore, if Weak RTM is to have any plausibility at all, it must be regarded as committed only to the claim

7. Stich, From Folk Psychology to Cognitive Science, op. cit., p. 186.

that *narrow* semantic properties supervene on syntactic properties. This will be more fully discussed later.

The difference between Strong and Weak RTM is, then, this: According to Strong RTM, the semantic properties of mental states are (i.e. are identical with) causally relevant properties. According to Weak RTM, the semantic properties of mental states are not causally relevant properties. It is the formal or syntactic properties of mental states which, for Weak RTM, are causally relevant. But these syntactic properties have semantic properties superveniently correlated with them. Therefore, according to Weak RTM, cognitive science is subject to what I shall call the *Syntactic Requirement*. One might seek to state this requirement thus: Cognitive psychology seeks to develop laws, generalizations, and explanations which relate mental state types in virtue of their syntactic properties. However, as indicated in Chapter I, the 'in virtue of' clause can prove troublesome in contexts such as this. For it leaves open the possibility of two non-equivalent interpretations of the above claim. On the one hand it might mean that cognitive psychology seeks explanations which are ultimately made possible by the syntactic properties of mental states, for these properties are the loci of causal *power*. On the other hand, it might mean that cognitive psychology seeks explanations which contain terms that name syntactic properties, for these properties are the loci of causal *relevance*. That is, the phrase 'in virtue of' can be given an ontological/power interpretation, or it can be given an explanatory/relevance interpretation. Problems arise when these two interpretations are conflated. However, since it is a version of the

principle of *explanatory* methodological solipsism which is being discussed, the appropriate interpretation is one which makes syntactic properties the loci of causal relevance. Thus a better statement of the syntactic requirement is this:

(S) Cognitive psychology seeks to develop laws, generalizations, and explanations which relate the syntactic properties of mental states, and which relate mental states derivatively upon relating their syntactic properties.

This claim will constitute the focus of the forthcoming discussion.

According to Stich's classificatory schema, along with Strong and Weak forms of RTM, there is also the Syntactic Theory of Mind (STM). Stich characterizes this as follows:

" The basic idea of the STM is that the cognitive states whose interaction is (in part) responsible for behaviour can be systematically mapped to abstract syntactic objects in such a way that causal interactions among cognitive states, as well as causal links with stimuli and behavioural events, can be described in terms of the syntactic properties and relations of the abstract objects to which the cognitive states are mapped. More briefly, the idea is that causal relations among cognitive states mirror formal relations among syntactic objects. If this is right, then it will be natural to view cognitive state tokens as tokens of abstract syntactic objects."⁸

According to Stich, STM views mental states as relations to purely syntactic objects (object types). There is some confusion as to what

8. From Folk Psychology to Cognitive Science, op. cit., p. 149.

this means. Firstly, STM need not be seen as committed to the implausible claim that ordinary folk-psychological propositional mental states do not have semantic properties. That is, unless it also wanted to be committed to some form of eliminativism, STM need not deny that mental states as taxonomized by folk psychology have semantic properties of some kind. Therefore, the claim of STM must probably be interpreted as follows: *For the purposes of cognitive science* mental states are to be type-individuated purely by reference to their essential syntactic type. For the purposes of cognitive science, the semantic properties of mental states play no part in their type-individuation. I will refer to this requirement as the *syntactic individuation*, or SI, requirement. The SI requirement is a precondition of the syntactic requirement.

Thus, according to STM, cognitive psychology is certainly committed to the syntactic requirement; it will seek laws, generalizations, and explanations which relate the syntactic properties of mental states. Thus Stich writes:

" The core idea of the STM - the idea that makes it syntactic - is that the generalizations detailing causal relations among hypothesized neurological states are to be specified indirectly via the formal relations among the syntactic objects to which the neurological state types are mapped. Similarly, generalizations specifying causal relations between stimuli and neurological states will identify the neurological states not by adverting to their essential neurological types, but, rather, by adverting to the syntactic objects to which the neurological types are mapped. Ditto for generalizations specifying causal relations between neurological states and behaviour."⁹

9. From Folk Psychology to Cognitive Science, op. cit., p. 151.

However, it is difficult to see how this can be the "core idea" of STM, since this claim in no way distinguishes it from Weak RTM. The claim is merely a statement of the syntactic requirement which applies equally to Weak RTM. Stich tries to reinforce the distinction with the further claim:

" On the matter of content or semantic properties, the STM is officially agnostic. It does not insist that syntactic state types have no content, nor does it insist that tokens of syntactic state types have no content. It is simply silent on the whole matter. But to put the point this way is perhaps a bit misleading, for in remaining agnostic on questions of content, the STM is in effect claiming that psychological theories have no need to postulate content or other semantic properties, like truth-conditions. It sees no psychological work to be done by the hypothesis that mental state tokens or types have semantic properties."¹⁰

Let us first purge Stich's suggestion of a few inaccuracies. To begin with, consider what Stich must mean when he says that STM does not claim "... that syntactic state types have no content." If by "syntactic state type" he means syntactic property, then the claim is trivial. Syntactic properties are not the sort of thing that can *have* content. It is mental states which have content, and syntactic properties are properties of mental states. Therefore, by "syntactic state type" Stich must mean something like "mental state individuated according to its essential syntactic properties." And Stich's suggestion must read something like this: STM is agnostic on the question of whether mental states, when type individuated in accordance

10. From Folk Psychology to Cognitive Science, op. cit., pp. 185-86.

with their essential syntactic properties, have semantic content. Rather, it sees "no psychological work" to be done by the hypothesis that syntactically type-individuated mental states have semantic content.

This, then, provides the distinction between STM and Weak RTM. Weak RTM is committed to the claim that semantic properties supervene on, or are at least associated with, syntactic properties. Hence, for Weak RTM, mental states individuated in accordance with their essential syntactic type necessarily have semantic content also. In contrast, STM is committed to no such supervenience claim. On the contrary, it sees 'no psychological work' to be done by the hypothesis that syntactically type-individuated mental states have associated semantic content. Thus, for STM, psychological theories have no need to postulate semantic content of any sort.

However, as a way of distinguishing Weak RTM and STM, this, by itself, simply will not do. By the claim that postulation of semantic properties does "no psychological work", Stich presumably means that nothing of relevance to a cognitive psychology turns on such postulation. But the same claim can be made for Weak RTM. For even though Weak RTM is committed to the supervenience claim, it need not see any useful psychological work to be done by the postulation of semantic properties. That is, whilst it may regard the supervenience claim as a true metaphysical fact, it may regard it as a fact of no relevance to psychology. To make this clearer, consider Stich's arguments against Weak RTM.

The essence of Stich's arguments against Weak RTM is that the postulation of semantic properties correlated with syntactic properties is, for the purposes of psychological theorizing, counterproductive.¹¹ He argues that the postulation of semantic properties places various constraints on psychological theory construction which "... impedes the quest for developmental, comparative, and clinical cognitive theories".¹² Stich's reasoning, here, appears to be as follows: Since Weak RTM must assume a supervenience relation between syntactic properties and semantic properties, any psychology which develops generalizations and explanations relating mental states in virtue of their syntactic properties must also be constrained by the relations between contents which these generalizations and explanations entail. However, Stich's reasoning, here, appears to be unfounded.

In particular, why must psychology, in the Weak RTM mould, be constrained by relations between contents? To be thus constrained, Weak RTM must regard psychology as having *two* essential concerns:

(i) To develop laws, generalizations, and explanations which relate mental state types in virtue of their syntactic properties (their essential syntactic types). And,

11. Stich, From Folk Psychology to Cognitive Science, op. cit., pp. 135-48.

12. Stich, From Folk Psychology to Cognitive Science, op. cit., pp. 207-208.

(ii) To explicate the supervenience relations obtaining between the syntactic properties of mental states and their associated (supervenient) semantic properties.

Only if one adopts (ii) as well as (i) can one view psychology (in the Weak RTM mould) as being constrained by relations between semantic contents.

However, it is possible to adopt Weak RTM without adopting claim (ii). Weak RTM is simply the conjunction of the syntactic requirement (S), the claim that cognitive science seeks laws, generalizations, and explanations which relate mental state types in virtue of their syntactic properties, with the claim that semantic properties are superveniently correlated with those syntactic properties. There is nothing in Weak RTM which commits it to claim (ii) above.

This is what is meant by saying that Weak RTM can regard the supervenience of semantic on syntactic properties as a metaphysical fact without regarding it as a fact which has any bearing on the formulation of psychological theories. Perhaps the supervenient semantic properties are unknown, or unknowable. This point necessitates a reformulation of the distinction between STM and Weak RTM. While both theories are committed to the claim that the explanations of psychology relate mental properties in virtue of their syntactic types, Weak RTM makes the further claim that semantic properties supervene on these syntactic properties. STM is agnostic with regard to this latter claim. However, neither STM nor Weak RTM need view the supervenience claim as a fact

relevant to psychology. If the above arguments are correct, neither STM nor Weak RTM need be constrained by relations between semantic contents. Weak RTM would entail such constraint only if it adopted the claim expressed in (ii) above. But this claim need be no part of Weak RTM.

This point assumes importance, of course, if the claim that semantic properties supervene on syntactic properties is true. And I shall go on to argue that a version of this statement is true. That is, I shall go on to claim that *narrow* semantic properties supervene on syntactic properties.

Now, if this principle is true, then STM's agnosticism with regard to (at least narrow) semantic properties cannot be sustained. Hence, STM would have to hold not only the syntactic requirement (S), but also the claim that narrow semantic properties supervene on syntactic properties. And this means that STM collapses into Weak RTM (reinterpreted to restrict the supervenience claim to *narrow* semantic properties). And it would not do to try and distinguish STM and Weak RTM by way of the claim that, for STM, no psychological work is done by the postulation of semantic properties; for this claim can be equally true of Weak RTM.

Due to the problems in adequately distinguishing STM from Weak RTM, I shall argue against the principle common to both of these, the syntactic requirement (S):

- (S) Cognitive science seeks to develop laws, generalizations, and explanations which relate the syntactic properties of mental states, and which relate mental states derivatively upon relating their syntactic properties.

Concomitantly, I want to argue that psychological theorizing is constrained by relations between semantic contents. To this end I want to establish the following pair of claims:

- (A) Narrow semantic properties are strongly supervenient on syntactic properties.
- (B) Narrow semantic properties are anomalous.

The bulk of the remainder of this chapter will be concerned with establishing these two principles. I will begin with (A); the claim that narrow semantic properties are strongly supervenient on syntactic properties. The first step in establishing (A) is to establish the following principle:

- (A₁) Syntactic properties, at least as they found in the philosophy of mind, are *functional properties*.

The following section will largely be concerned with defence of the claim expressed in A₁. This defence will consist in an examination of the concept of syntax, as it appears both in linguistics, and in the philosophy of mind.

4

A useful place to begin discussion of the notion of a syntactic property is with a quotation from Fodor:

" 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. Since we don't know how to complete this list (since, that is, we don't know what semantic properties there are), I see no responsible way of saying what, in general, formality amounts to. The notion of formality will thus have to remain intuitive and metaphoric, at least for present purposes: formal operations apply in terms of the, as it were, *shapes* of the objects in their domains."¹³

Now Fodor, at least on one interpretation, defends Weak RTM; he defends a view of cognitive science as seeking generalizations and explanations which apply to mental states in virtue of their syntax. One begins to wonder why if his grip on the notion of syntax is

13. "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology", op. cit. Reprinted in representations, pp. 225-53, p. 227.

"intuitive and metaphoric". Let us see what can be gleaned from the passage.

Firstly, for Fodor, formal properties and syntactic properties are not identical ("... since we could have formal processes defined over representations which don't in any obvious sense have a syntax. Rotating an image would be a timely example.").¹⁴ Rather, the relation is this: if a property is syntactic then it is also formal, but not vice versa. And this relation means that the intuitive and metaphoric character of formal properties also applies to syntactic properties. Thus, in the quoted passage, the only grip Fodor appears to have on the notion of a syntactic property is that, "... being syntactic is a way of not being semantic". The only positive suggestion put forward by Fodor is that "... formal operations apply in terms of the shapes of the objects in their domains." I will venture, then, the following interpretation of Fodor's positive claim: The syntactic properties possessed by a given mental state are (identical with) the properties of that state of being such and such a *shape*. The syntax of a psychological state is the *shape* of that state. This is the view I shall argue against.

It might be interesting to see what Fodor's old colleague Jerrold Katz would think of the claim that syntactic properties are properties of shape if that claim had been made in the sphere of

14. "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology", op. cit. Reprinted in Representations, op. cit., pp. 225-53, p. 227.

linguistics. For, in that sphere, the claim is a hopelessly inadequate characterization of the concept of syntax.

The notion of syntax is intimately connected with the notion of *structure*; the syntax of a sentence is, in some sense, the structure of that sentence. However, there is more to structure than merely shape. Consider the notion of sentence structure. Clearly the structure of a sentence depends on such things as the order of words occurring in that sentence - to this extent it depends on the 'shape' of the sentence. However, it also depends on much more.

Many sentences in a language are *structurally ambiguous*. They possess ambiguities that are additional to any lexical ambiguities that they may contain. For example:

Dafydd likes exciting sheep.

This sentence is ambiguous in a way that cannot be traced to the ambiguity of any words it contains. It might mean, on the one hand, that Dafydd enjoys causing sheep to become excited. On the other hand it might mean that he enjoys the company of interesting and stimulating sheep.

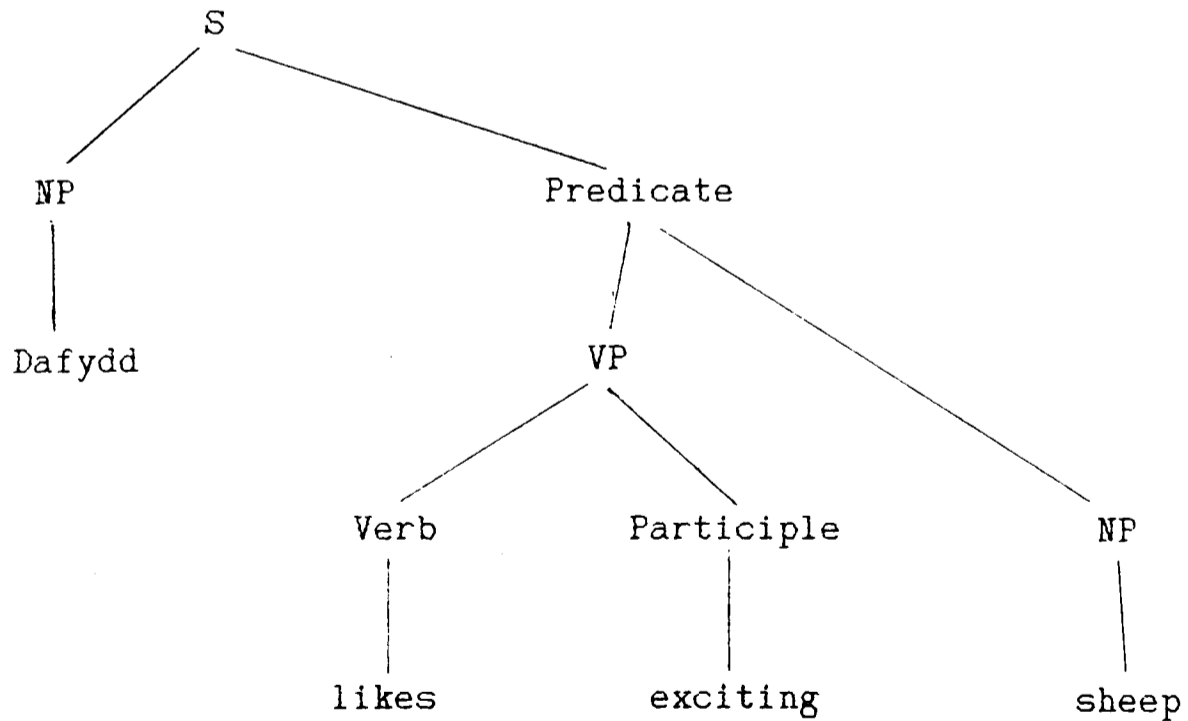
The ambiguity arises because the sentence 'Dafydd likes exciting sheep' has two distinct structures. One will group 'likes' and 'exciting' together as a single constituent of the sentence, the other will group 'exciting' and 'sheep' together as a constituent. This shows that there is sometimes more than one way of assembling words into subsentential components of a sentence. There is more to the structure

of a sentence than simply the order of words out of which the sentence is built. That is, there is more to the syntax of a sentence than its 'shape'. There are intermediate layers of sentence organization between the level of words and the level of the sentence itself.

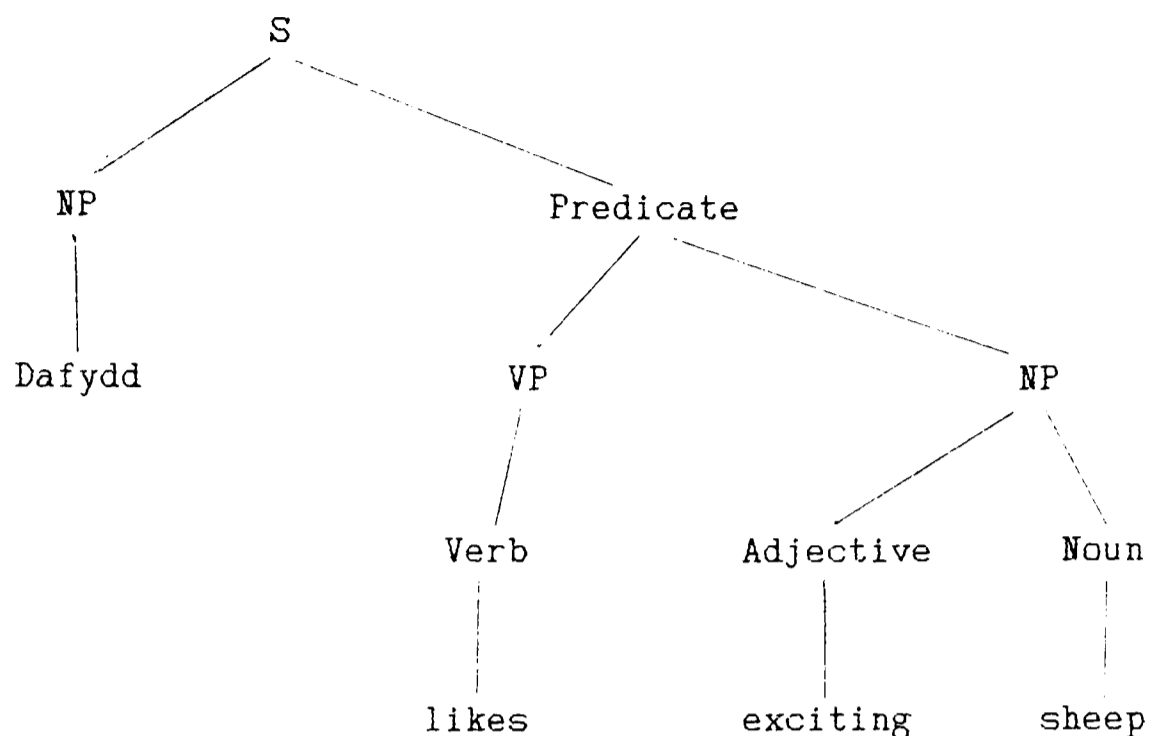
It is common to represent the internal structure of sentences as trees which display the hierarchical organization of sentences. If the sentence is structurally ambiguous, it gets more than one tree. These trees are called *phrase-structure* trees because they organize the sentence elements into successively larger constituents, or phrases, of the sentence.

The two trees for 'Dafydd likes exciting sheep' are as follows:

Dafydd 1.



Dafydd 2.



Both Dafydd 1 and Dafydd 2 are oversimplified. However, they are right in dividing S (the sentence) into two basic units, NP (noun phrase) and Predicate. Dafydd 1 differs from Dafydd 2 only in the structure of its predicate. In Dafydd 1, the predicate is composed of a complex VP (verb phrase) and simple NP. In Dafydd 2, the reverse is the case. As a result of these different underlying structures, the sentence has two sets of truth-conditions and, hence, two distinct meanings.

The crucial point here is that the predicate 'likes exciting sheep' is syntactically different in each case. And the reason for this syntactic difference is that the predicate *functions* differently in each case. The lexical items comprising the predicate enter into different relations of combination with one another in each case. Therefore, what determines the syntax of each lexical item, is not the 'shape' of those

items, but it is the way they combine with one another to form a predicate. Similarly, what determines the syntax of the predicate is not the 'shape' of the predicate; it is not, for example, the mere order of its lexical constituents. Rather, the determinants of syntax, here, are the relations of combination into which the lexical constituents enter. In linguistics, syntactic properties are properly understood to be *combinatorial* properties: a given syntactic property is individuated by the relations it enters into with other syntactic properties, and whether or not some item possesses a given syntactic property is determined by the relations that item enters into with other items which have certain other specified syntactic properties.

The equation of syntactic properties with combinatorial properties holds whether one believes the locus of genuine syntax to be in surface structure *or* in deep structure.

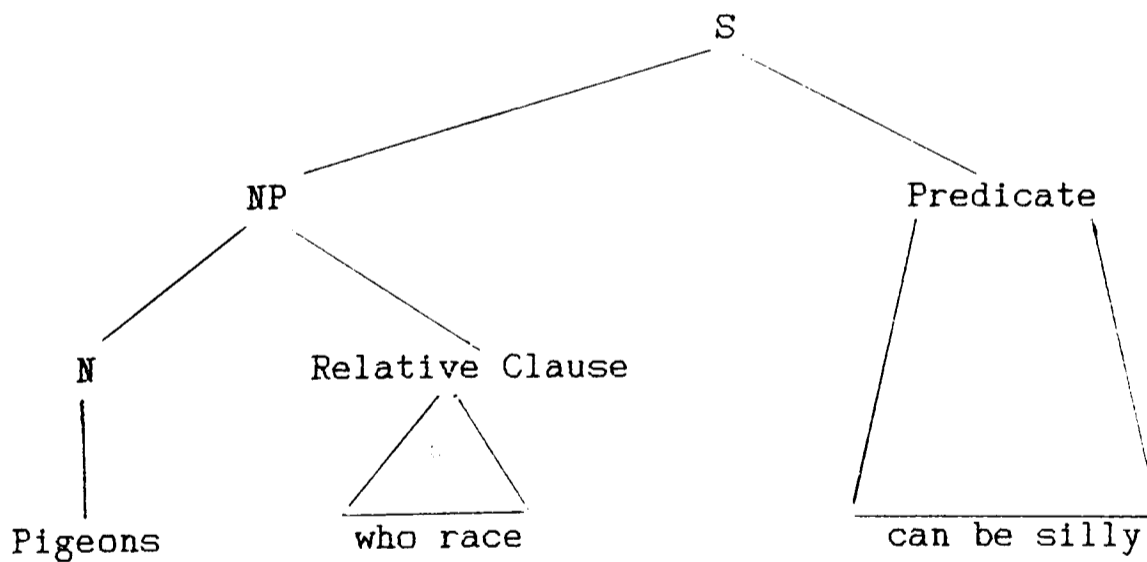
The notion of deep structure is a notion associated with Noam Chomsky.¹⁵ Chomsky claims that we need to view a sentence as having two structures, both represented as trees. The two structures are linked. One of these he calls the *deep* (or *base*) structure. The other, discussed above, he calls the *surface* structure. Consider, for example, the following sentence:

Racing pigeons can be silly.

15. See, for example, N. Chomsky: Syntactic Structures (The Hague, Mouton, 1957); Aspects of the Theory of Syntax (Cambridge, MIT Press, 1965); Topics in the Theory of Generative Grammar (The Hague, Mouton, 1966).

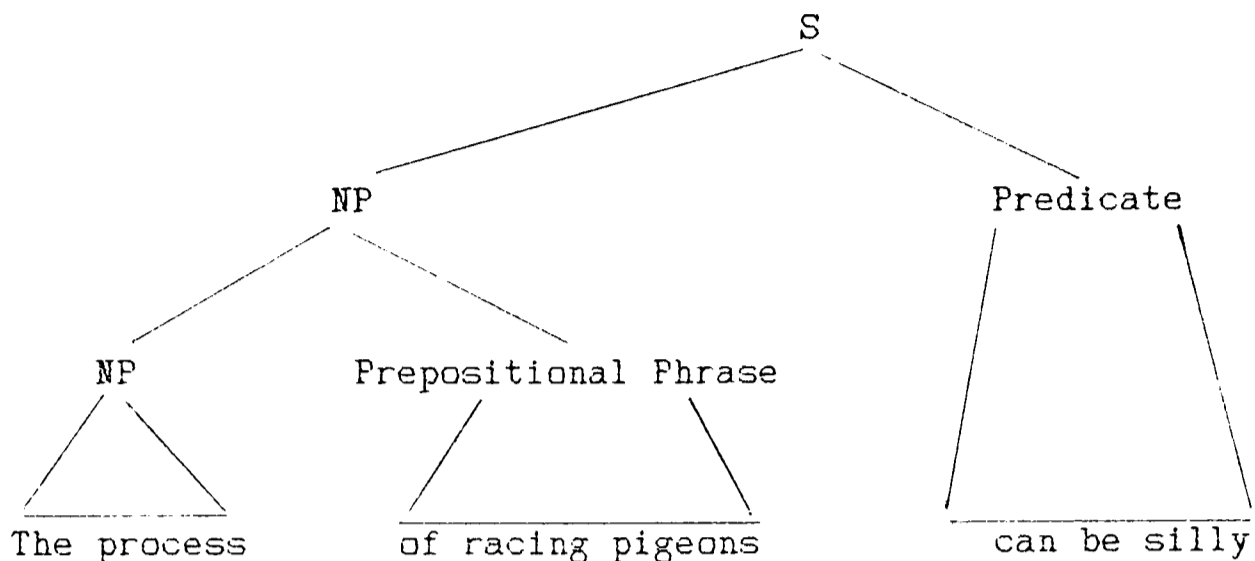
This is structurally ambiguous. It might mean that either that it is silly to race pigeons, or that pigeons who race can be silly. However, this ambiguity is quite different from the type considered above. There the ambiguity was a result of different ways of grouping words into constituents of the sentence. Here it arises *within* the constituent 'racing pigeons'. It is as though that phrase can go proxy for two distinct phrases: 'the process of racing pigeons' and 'pigeons who race'.¹⁶ Chomsky argued that to account for structural ambiguities such as these, we need to postulate distinct *deep* structures underlying the ambiguous sentence. Oversimplifying, the alternative deep structures would be something like this:

Pigeon 1



16. Both the 'exciting sheep' and the 'racing pigeons' examples are based on examples given by Michael Devitt and Kim Sterelny, Language and Reality (Oxford, Basil Blackwell, 1987), Ch.6, especially pp. 92-97.

Pigeon 2



Transformational rules link both deep structures to a single surface structure by deleting and moving elements in the sentence.

Once again, the crucial point is that the noun phrase 'racing pigeons' has two distinct syntactic forms. And this has nothing to do with the 'shape' of that phrase; for on any reasonable construal of 'shape', the shape of the phrase remains constant. Rather, the noun phrase has two distinct syntactic forms because it enters into different combinatorial relations in each case. It is the relations which the phrase enters into with other subsentential and sentential items that is the determinant of the syntax of that phrase, and not any supposed 'shape' that the phrase possesses.

There is a tendency to identify syntactic properties with one or another physical property of natural language tokens. The view of syntax as shape is one, perhaps the most plausible, manifestation of

this tendency. However, the arguments given above, if correct, show this approach to be misconceived.

The syntactic properties of phrases such as 'likes exciting sheep' or 'racing pigeons' are determined not by any physical features of the written tokens of those phrases, but rather, by the relations of combination which those tokens can enter into with other linguistic items, lexical, subsentential, and sentential. And this point applies quite generally throughout natural language. There is no physical property, possessed by any token linguistic item, which can determine the syntax of that item. Within certain parameters, differences in the shape of written tokens do not count as syntactic differences. Intuitively, 'S' can be syntactically identical with '8' or with '9' etc., etc. These tokens could not count as syntactically distinct because that would mean that a given letter could have as many syntaxes as there were token-instantiations of it. Therefore, syntactically type-identical tokens can have different shapes: identity of shape is not necessary for syntactic identity. Moreover, it is not at all clear that the parameters of allowable variation in shape are in any sense narrow. For example, morse code tokens occurring in time can share their syntax with printed English tokens.¹⁷ Thus, 'S' and '...' are syntactically type-identical.

17. This point is made by Dennett in, "Beyond Belief", op. cit., p. 21.

These arguments suffice to show that possession by a linguistic item of the property of having such and such a shape is not necessary for that item's possession of any given syntactic property. Similar remarks apply to any physical property whatsoever. What physical properties do 'S' and '...' have in common which would be relevant to determining their syntax?¹⁸ In linguistics, syntactic properties are not physical properties, they are *combinatorial* properties. Any token linguistic item, as a physical token, may possess any number of physical properties, but what determines the syntax of that token is not any of those physical properties but the relations of grammatical combination into which the token may enter with other tokens. Another way of putting this is to say that syntactic properties can be *variably realized* physically. Combinatorial roles can be implemented in diverse physical ways, and so the syntactic properties of natural language abstract across physical properties.

Conclusion: The syntactic properties of natural language are not physical properties. They are combinatorial properties.

Similar remarks apply not only to the concept of syntax in linguistics, but also to that concept in the philosophy of mind. The problem here is of delineating the sort of identity conditions

18. Dennett also makes the interesting point that the only features of physical object-tokens we would regard as syntactic are those which make, or realize, a semantic difference., I think that Dennett here is exactly right. This is precisely the line taken in this Chapter. See, "Beyond Belief", op. cit., p. 21 and p. 23.

applicable to syntactic properties not of natural language but of the so-called *language of thought*.¹⁹

First, a brief note on my use of the notion of a language of thought. I do not take the term 'language' very seriously here. I assume that thoughts and processes of thinking are token-identical with neurophysiological state and process tokens. And due to this it is appropriate to speak of thoughts and thought processes being expressed by some neurophysiological code. And the neurophysiological code can then, in some sense, be said to constitute the language of thought. However, this is all the expression 'language of thought', as I use that expression, is meant to imply. Consequently, this is as far as I take the 'little sentences in the head' hypothesis. It is *not* meant to imply that the neurophysiological code need share many features of the sentences of natural language. It is not meant to imply, for example, that the neurophysiological code has logical structure etc.²⁰

We can picture a language of thought as follows: The language is composed of mental representations. These are the analogues of sentences of natural language. Every representation has associated with it semantic properties; and these are a function of the

19. *Locus Classicus*, Jerry Fodor: The Language of Thought (New York, Thomas Y. Crowell, 1975). It will become clear that my notion of the language of thought is nowhere near the full Fodorian sense.

20. Pace, Fodor, The Language of Thought, op. cit., and Field, "Mental Representation", op. cit.

representation's representational (i.e. referential truth-conditional) properties and a function of its properties of functional role. Also associated with the representation are, as yet unspecified, syntactic properties. Mental representations are realized by neurophysiological states and processes (just as sentences of natural language are realized by strings of written or spoken inscriptions or utterances), and these states and processes thus form some sort of neurophysiological code (just as written inscriptions form a code).

The problem is this: not every aspect or property of the neurophysiological states or processes which realize any given mental representation can count as a syntactic property (just as the typeface, size, colour, and, if the foregoing arguments are correct, *shape* of written inscriptions do not count as syntactic properties). The problem, then, is identifying a subset of the properties of the neurophysiological states and processes which are syntactic properties. As Dennett puts it, the problem is distinguishing neurophysiological 'scribbles and smudges' from genuine syntax.²¹

In the case of natural language, the suggested subset consisted of the property of having shape. So, as applied to the language of thought, the suggestion would be that it is the property possessed by neurophysiological states and processes of having such and such a shape which constitute the syntactic properties of the language of thought.

21. The expression is Dennett's. "Beyond Belief", op. cit., p.25.

However, just as with natural language, the arguments from variable realization come into play: If syntactic property A can be realized by the property, possessed by a neurophysiological state, of having shape B in one instance, this does not ensure uniformity throughout all actual, or more to the point, *possible* cases. Therefore, the syntactic properties of the language of thought cannot be identified with the properties of having shapes of a certain kind.

Moreover, just as with natural language, the principle of variable realization applies across all physical properties. Whatever physical property of the various neurophysiological states and processes is chosen, the principle of variable realization undermines the attempt to identify a syntactic property with that physical property.

Conclusion: The syntactic properties of the language of thought are not physical properties.

The notion of syntax can, in this regard, be usefully compared with that of valence.²² The notion of valence is one associated with atoms. An atom of an element or compound is composed of an atomic nucleus and orbiting rings of electrons. The atomic nucleus is made up of protons and (except in the case of the protium isotope of hydrogen) neutrons. The electrons travel around the nucleus in various orbits which are statistically precise and unexceptional. Electrons are never,

22. Dennett talks of syntactic properties as marking 'combinatorial valences'; "Beyond Belief", op. cit., p. 21. However, he does not take the analogy any further.

statistically, found in between orbits. An atom may be said to have 'shape' (in some sense) in virtue of the configuration of protons, neutrons, and electrons which constitute it.

However, it is not in virtue of every aspect of this shape that the atom enters into physical interactions with other atoms. Largely determining this sort of interaction is the number of electrons in the outer electron orbit. The valence of an atom is the combinatorial possibilities of that atom, and the assignment of integers as the values of valences corresponds to the number of electrons in the outer electron shell. Thus, if an atom has two electrons in its outer electron shell, it will have a valence of two. Therefore, the valence value of the atom corresponds to the number of electrons in the outer electron shell of that atom. Does this mean that valence can be identified with outer electron number? More precisely, does this mean that the property of having valence V can be identified with the property of having V number of electrons in the outer electron orbit? If so, valence would appear to be a physical property.

However, in fact, the property of having a given valence cannot be identified with any physical property: the property of having valence V cannot be identified with the property of having V number of electrons in the outer electron shell. The property of having a given valence abstracts across physical structures. It is not a physical property, as that notion is commonly understood.²³

The reason for this is, once again, the possibility of variable realization. For example, the identification of valence with outer electron number does not hold in the case of covalent bonding. In such bonding, combinatorial possibility is not determined by the number of electrons in the outer electron shell. Moreover, when atoms get very large, their nucleic mass becomes a more important determinant of combinatorial possibility than outer electron number.

One response to this would be to say that valence is a different property in each case.²⁴ But this would undermine any attempt to say what justifies our talking about valence, as a unitary phenomenon, at all.²⁵ We want to be able to say that Potassium has a valence of 1, that Carbon has a valence of 4, and that Uranium has a valence of 2, even though the valences are, in each case, realized by physically quite diverse means.

The crucial idea to grasp here is that valence is to be understood as a *combinatorial* property; it is to be specified in terms of a given atom's dispositions to interact with other atoms. Once such a specification has been completed, there is nothing more which can be said about what the valence of that atom is, though a great deal more

23. In accordance with Blackburn, "Losing Your Mind: Physics, Identity and Elimination" (unpublished), we might modify this to the claim that the property of valence is not a 'Tractarian' or constitutional physical property.

24. This response is modelled on the P.S. Churchland approach to physical properties such as temperature. See Chapter I.

25. See Chapter I, the discussion of *scientific explanatory properties*. Also n.16, Chapter I.

can be said about the way that valence is, physically, implemented in any particular instance.

Therefore, the property of having valence V is to be understood as a combinatorial or functional property: the valence of an atom is specified in terms of that atom's dispositions to interact with other atoms. When, and only when, this functional specification has been completed can we begin looking for structures - for example, outer electron number, nucleic mass - which will realize the function; structures, that is, in which the function is instantiated. But, for the familiar reasons stemming from variable realization, the functional property cannot be identified with the structural property which, in any particular instance, realizes it.

This point can be pushed further. Valence is a functional property. In each of its instantiations, however, it has a structural, i.e. physical, realization. That is, it is instantiated in a physical thing, and certain physical properties of that thing realize the valence in question. Thus, it is these realizing properties which are the physical properties with which valence has often been, incorrectly, identified. For example, the property of being a valence V can be realized by, a certain electronic configuration. It is this electronic configuration type which is the physical property with which many have sought to identify valence. But this property cannot be identified with the valence on pain of contradicting the principle of variable realization.

I want to claim that syntactic properties are similar to properties of valence. Syntactic properties are individuated functionally. For any token linguistic item (word, letter, phrase, sentence) of a language - either natural language or the language of thought - the syntactic properties of that token are the dispositions that token possesses to interact with other linguistic tokens. The identity conditions of any syntactic property are determined by the dispositions which it bestows on tokens which have it to interact with other tokens, where this interaction takes place in virtue of the syntactic types of those tokens. A syntactic property is thus to be understood as a functional property; a property with functional identity conditions. Of course, a functional property such as this may, in any given language or group of languages, be realized by the same *shape* - in the sense of being realized in each particular instance by the same physical property. And, indeed, in each particular instance this physical property may determine the functional role of the linguistic token, that is, it may determine the syntax of the token. And this tempts the identification of the syntax with the physical property. But this identification rests on a conflation of a property with its realizing base property. And such identification would be as wrong as the identification of valence with number of outer electrons of an atom.

With particular application to the language of thought: Those properties of neurophysiological states and processes which we regard as syntactic are functional properties. The only way we can identify genuine syntactic properties, as opposed to neurophysiological scribbles and smudges, is in terms of functional role: genuine syntactic

properties are those properties which determine interactions between elements of the neurophysiological code. Of course, these functional syntactic properties are physically realized. And these physical realizations may, in any particular instance, be determinants of the functional role. It may thus be tempting to identify the syntactic property with the physical realization. But this would be no more correct than identifying valence with outer electron number.

Conclusion: The syntactic properties of both natural languages and the language of thought are combinatorial or functional properties.

This idea needs a little sharpening up. This will proceed by way of a distinction between combinatorial properties and functional properties. Until now, the two notions have been understood as identical. The distinction to be drawn now reflects an important difference between the notion of syntax in linguistics and that notion in the philosophy of mind.

The difference is this. Psychological syntax, by its very nature, is conceived of as having causal power/relevance, linguistic syntax is not. Psychological syntax is essentially viewed as something which plays a causal role in the bringing about of behaviour. The same cannot be said of linguistic syntax. For example, the following passage from Fodor expresses what I believe to be a fairly standard conception of the nature of syntax in psychology:

" Because, to all intents and purposes, syntax reduces to shape, and because the shape of a symbol is a potential determinant of its

causal role, it is fairly easy to see how there could be environments in which the causal role of a symbol correlates with its syntax. It's easy, that is to say, to imagine symbol tokens interacting causally *in virtue of* their syntactic structures. The syntax of a symbol might determine the causes and effects of its tokenings in much the same way that the geometry of a key determines which locks it will open."²⁶

Furthermore, the principle of formal explanatory methodological solipsism is simply a statement of the claim that syntactic properties in psychology have causal relevance.

This important difference between psychological and linguistic syntax will be recorded in a distinction between functional and combinatorial properties. In the psychological case, syntactic properties can be viewed as straightforward functional properties in the commonly understood sense of that term. That is, psychological syntactic properties are individuated by way of their *causal* roles. Linguistic syntactic properties, on the other hand, are combinatorial properties in that they are individuated by their mutual interactions, where, however, these interactions are not causal in nature.

In the following section I will argue that, in psychology, narrow semantic properties supervene on syntactic properties. This argument will depend essentially on the claim that syntactic properties are functional properties in the causal sense. Consequently, the argument does nothing to establish the corresponding claim for linguistics.

26. Psychosemantics, op. cit., pp. 18-19.

All the materials are now at hand to argue for the claim that, in psychology, narrow semantic properties supervene on syntactic properties. The argument trades on the fact that both narrow semantic properties and syntactic properties have constitutive links with behaviour. Firstly, narrow semantic content was introduced as that aspect of semantic content which plays a causal/explanatory role in the production of behaviour. So the constitutive links between narrow content and behaviour cannot plausibly be denied. Narrow contents are type-individuated in accordance with their causal/explanatory role. Secondly, given that syntactic properties are functional properties, where functionalism here implies causality, it seems that they, also, must have essential links with behaviour: syntactic properties are type-individuated in accordance with the role they play in bringing about other syntactic properties, and, ultimately, in the bringing about of behaviour. Given these two facts, the following argument can be constructed for the supervenience of narrow semantic properties on syntactic properties.²⁷

Firstly, given the constitutive links narrow content has with behaviour, it appears that a difference in narrow content (i.e. narrow semantic properties), uncompensated for by other differences, implies a

27. This argument is a suitably modified version of McGinn's argument for the claim that mental properties supervene on physical properties. The Character of Mind, op. cit., p. 29.

difference in behavioural dispositions. Secondly, it also seems undeniable, given that syntactic properties are to be understood as functional properties, that a difference in syntactic properties, uncompensated for by other differences, implies a difference in behavioural dispositions. That is, it seems undeniable that a difference in the behavioural dispositions of two organisms depends, other things being equal upon a difference of syntactic structures occurrent in their brains; for given that the arguments for the functional individuation of syntactic properties are correct, the only grip we have on the notion of a syntactic property is the role it plays in bringing about other syntactic properties, and, in conjunction with these, behaviour.

Suppose, then, that two organisms exemplified totally different narrow semantic properties, but also that they exemplified identical syntactic properties. This would imply that their dispositions to behaviour were correspondingly different, even though they were syntactically type-identical. And this result is plainly unacceptable. For it undermines the constitutive links holding between syntactic properties and behaviour.

The solution to this problem is to rule out the sort of independent variability of narrow semantic and syntactic properties which gives rise to the above dilemma. And this can be done by accepting some sort of correlation thesis between the two types of properties. Obviously, this correlation cannot be biconditional in character - this would violate variable realization. Therefore, the correlation must be of a

conditional character. That is, narrow semantic properties must *supervene* on syntactic properties.

The above argument, if correct, suffices to show that narrow semantic properties *supervene* on syntactic properties. However, nothing so far has been given to show that such supervenience must be *strong* rather than *weak*. However, if arguments stated in chapter I are correct, such supervenience must be strong. It was part of the burden of chapter I to show that the major difficulty with *weak* supervenience was not that it stated a dependence relation which was too weak to be used in discussing relations between mental and physical properties. That is, the problem with weak supervenience was not that it specified a dependence relation but one that was inadequate for talking of relations between mental and physical properties. Rather, the problem was that weak supervenience failed to specify an adequate dependence relation *of any sort*. The relation expressed by weak supervenience is just not a (coherent) relation of dependence. If this is correct, then, by elimination, the relation between narrow semantic and syntactic properties must be one of *strong* supervenience.

Conclusion: Narrow semantic properties are strongly supervenient on (psychological) syntactic properties.

In this section I will argue for claim (B): Narrow semantic properties are anomalous with respect to syntactic properties. The argument here will much quicker, as this claim is much more likely to be accepted.

The first premis in this argument is the claim that ordinary wide semantic properties are anomalous with respect to their syntactic bases. A proper defence of this claim would, perhaps, involve, another complete thesis. I will assume that such a defence can be mounted. The claim is likely to be accepted by many, though not by all. A major source of this claim is likely to be Quine's thesis of the indeterminacy of translation.²⁸ As formulated, this only establishes the anomalism of semantic properties relative to physical bases, but the required modification would presumably be a fairly simple matter. Quine's thesis, however, is not the only source of the claim.

McGinn argues that some sort of neo-Davidsonian argument could be run as follows: Relations between semantic contents are essentially governed by their own distinctive principles (interlocking with the psychological) and these principles have 'no echo' in physical theory. Semantico-physical laws would transmit individuation or attribution conditions thus undermining the above principles and,

28. W.V.O. Quine: Word and Object (Massachusetts, MIT Press, 1962), especially ch.2.

effectively, 'changing the subject'.²⁹ Again McGinn uses this argument primarily to establish the anomalism of semantic properties relative to the physical. But again, the required modification necessary to rule out semantico-syntactic laws would be very easy to make. Indeed, since the syntactic properties of psychology are functional properties, this modification would amount to no more than the claim that there can be no non-intentional, specifically functional, interpretation of relations between semantic contents. Such a claim was discussed and defended at length in Appendix III.

The argument given here is brief as, I believe, most people would be antecedently prepared to accept the anomalism of ordinary wide semantic properties.

Given that ordinary semantic properties are anomalous, the claim that narrow semantic properties are also anomalous is strongly supported. The reason for this stems from the claim, established earlier, that narrow semantic properties cannot be abstracted from ordinary wide semantic properties in such a way as to be capable of independent existence or expression. The only way we have of specifying narrow semantic properties, consequently, the only grip we have on the notion of a narrow semantic property, is by way of ordinary wide semantic properties. But if narrow semantic properties are not capable of sustaining such abstraction, then they do not seem to be the sorts of

29. "Philosophical Materialism", op. cit., p. 194.

things which, *by themselves*, can enter into nomological relations with syntactic, or physical, properties. How could they? What sorts of lawlike statements could express such nomological relations? The discussion of Chapter II should have made it clear that there can be no such statements.

Of course, if ordinary wide semantic properties were nomological, then there would be a sense in which narrow semantic properties were nomological also. This nomologicality would be inherited from the ordinary semantic properties. Given that narrow semantic properties cannot be abstracted from ordinary semantic properties, they *must* inherit this nomologicality. However, ordinary semantic properties are not nomological, they are anomalous. And, in this case, similar remarks apply. Given the essentially dependent nature of narrow semantic properties, they inherit the anomalism of wide semantic properties.

Conclusion: Narrow semantic properties are anomalous with respect to their syntactic bases.

7

The discussion of the previous two sections has been concerned with establishing the following principles, which, given the correctness of the arguments employed, can now be accepted:

(A) Narrow semantic properties are strongly supervenient on syntactic properties.

(B) Narrow semantic properties are anomalous.

Once again we have a prima facie inconsistency. This apparent inconsistency has been met several times in this paper, and each time has been resolved in accordance with the following principle.

(α) If property M is (a) anomalous, and (b) strongly supervenient on supervenience base property S composed of realizing base properties $R_1 \dots R_n$ then, (i) S has no natural specification independently of M, and (ii) none of $R_1 \dots R_n$ have a natural specification independent of M.

Thus in the case of wide mental states, for example, it was argued that the only natural specification of the identity conditions of the realizing bases of these states must make essential reference to the wide mental states which they realized. Similarly with the case of internal components vis-a-vis their realizing bases. It was argued that the anomalism of internal components could be reconciled with their strong supervenience on the physical by way of the claim that the only natural specification of the identity conditions of the realizing base of any given internal component involved essential reference to the internal component which the base property realized.

I advocate the same step in the case of narrow semantic properties vis-a-vis their syntactic realizations. That is, I advocate we make the following claim:

- (C) Any given narrow semantic property, attaching to a mental state is realized by a syntactic property. However necessarily, any natural specification of the identity conditions of this syntactic property must make essential reference to the narrow semantic property which it realizes. Hence, the syntactic property does not have purely syntactic identity conditions.

To see how this condition might obtain, it will be useful to compare it to the case of wide mental states. Here it was argued that the realizing bases are made up of distinct physical properties which could only be unified, drawn together to form a single property, in a way which was not blatantly gerrymandered, if reference is made to the wide mental state which they together realize. In the case of narrow semantic properties and their syntactic realizations the situation is similar, and is probably something like the following: Syntactic properties are (functional) properties of neurophysiological states and processes which can only be differentiated, recognized as discrete and distinct properties, because of their capacity to make a difference to the narrow semantic properties possessed by an organism.

However, in this case there is a slight complication. For narrow semantic content cannot be *expressed* in sentences of natural language.

Reference to narrow semantic content, and hence to narrow semantic properties is accomplished by *mentioning* it, or *indexing* it with ordinary (wide) content sentences of natural language. Therefore, a more correct form of (C) would be:

(C') Any given narrow semantic property attaching to a mental state is realized by a syntactic property. However, necessarily, specification of the identity conditions of this syntactic property must make essential reference to the ordinary wide content sentence which indexes the narrow content which the syntactic property in question realizes.

The specification of the identity conditions of any given syntactic property must make essential reference to the wide content sentence which indexes the narrow semantic properties which the syntactic property in question realizes. Therefore, generally, syntactic properties can only be individuated in terms of the narrow semantic properties they realize, and this means that they can only be individuated in terms of the ordinary wide content sentences which index the narrow semantic content that they realize.

What I have called the *syntactic requirement* (S) is stated as follows:

(S) Cognitive psychology seeks to develop laws, generalizations, and explanations which relate the syntactic properties of mental

states and which relate mental states in virtue of relating their syntactic properties.

Given the correctness of the foregoing remarks, this principle must be rejected. A precondition of (S) acting as a constraint on the formation of scientific psychological theories is the syntactic individuation (SI) requirement. Mental states must be individuated in accordance with their essential syntactic type. The above arguments, if correct, show that this requirement cannot be satisfied. Syntactic properties can only be individuated in terms of the semantic difference they make. Hence, (S) must be rejected.

Conclusion.

The basis of this thesis has been an essentially deflationary approach to the constraints placed upon psychological theorizing by the principle of methodological solipsism.

In Chapter I, it was argued that the principle of methodological solipsism, viewed as a constraint on the formation of psychological theories, arises from what was described as a *causal objection* to the use, in those theories, of folk psychological locutions, and to the propositions or content sentences upon which they are based. The important distinction between causal *power* and causal *relevance* was defended, and, on the basis of this distinction, two importantly different forms of methodological solipsism were isolated. These were *ontological* methodological solipsism (OMS) and *explanatory* methodological solipsism (EMS). It is only EMS which provides any substantive methodological constraint. Accordingly, the remainder of the thesis was concerned with this principle. Within the category of EMS a further distinction was drawn. This was between *conceptual* explanatory solipsism (CEMS) and *formal* explanatory methodological solipsism (FEMS).

The subject of Chapter II was CEMS. It was argued that the explanatory notion of *narrow content* invoked by CEMS is an incoherent notion in so far as narrow semantic content can only be individuated in terms of the wider semantic content of which it forms a proper part. Narrow semantic content is individuation dependent on ordinary wide

semantic content. In Chapter II also, the beginnings of a positive account were sketched, an account which attempts to show how ordinary wide semantic content can feature in the explanations of a developed cognitive psychology even though it is not taxonomized according to causal power. This account was based around the notion of *indexing*. The rationale behind this was again the distinction between causal relevance and causal power, and the subsequent distinction between OMS and EMS. The guiding principle here was that ordinary semantic content can have causal relevance in the absence of causal power taxonomization. And this amounts to the claim that EMS can be rejected while OMS is accepted.

Chapter III began the long argument against FEMS. In Part 1 Davidson's arguments for the anomalism of mental properties were clarified and strengthened. It was argued that these must rest upon the impossibility of transmission, from physical to mental properties, of individuation conditions which are both *natural* and *physical*. Part 2 was concerned with motivations for, and characterization of, a mental-physical property supervenience thesis. Part 3 examined the question of the compatibility of the principles of anomalism and supervenience. It was argued that a mental property could strongly supervene on a physical property, consistently with the anomalism of the mental property only if a principle of the following form holds:

- (α) If property M is (a) anomalous, and (b) strongly supervenient on supervenience base property S composed of realizing base properties $R_1 \dots R_n$, then (i) S has no natural specification

independently of M, and (ii) none of $R_1 \dots R_n$ have a natural specification independently of M.

Chapter IV was concerned with clarifying and strengthening the conclusions of the previous Chapter by applying them to mental properties conceived within a dual component framework. Within this framework, one possible reconciliation of the principles of anomalism and supervenience was this: those properties of mental states constitutive of their internal components are strongly supervenient on natural physical states, but those properties constitutive of the external component are anomalous. This position was rejected on the grounds that it entailed an untenable account of the relation rationality bears to internal and external components. The position finally arrived at was that both internal and external components are strongly supervenient on physical properties, and both internal and external components are anomalous. This position was rendered consistent by way of the claim that the realizing bases of both internal and external components are not natural physical properties. This was a direct application of principle (α).

Appendix II was a defence of the reconciliation of anomalism and supervenience implicated in principle (α) against the claim that it could not account for the determination of the mental by the physical. Appendix III investigated some consequences the reconciliation underwritten by principle (α) has for the nature of explanation in psychology.

Chapter V was concerned with the application of the principles delineated in Chapters III and IV, in particular principle (α), to the principle of formal explanatory methodological solipsism. Two crucial claims were argued for: (a) narrow semantic properties are anomalous, (b) narrow semantic properties are strongly supervenient on syntactic properties. The only possible reconciliation of the claims expressed in (a) and (b) appeared to be by way of principle (α). And this entailed that there could be no natural specification of the individuation conditions of syntactic properties independently of the narrow semantic properties they realize. Given the correctness of the arguments of Chapter II, this must be modified to the claim that syntactic properties have no natural specification independently of the ordinary semantic content they (in part) realize. That is, syntactic properties are individuation dependent on ordinary wide semantic content.

Therefore, the rejection of both CEMS and FEMS are, ultimately, for similar reasons: the explanatory properties invoked by both principles are individuation dependent on ordinary semantic content. With the rejection of both CEMS and FEMS, the rejection of EMS is complete.

Just as all explanations must come to an end, so all arguments must have a beginning. And an essential premise in much of this argument has been that Davidson's case against psychophysical laws is largely correct. The justification for this assumption is twofold. On the one hand, a proper examination of this topic would require at least another thesis. On the other hand, Davidson's arguments are accepted by a large

number, almost certainly a majority, of people. However, the arguments given in Chapters III-V will probably be attractive only to such people.

The bulk of the thesis, then, has been essentially negative. What positive threads there are play a minor role in the thesis, and centre around the distinction between causal power and causal relevance, and the resulting notion of *indexing*. The claim that it is possible to accept OMS while rejecting EMS centres around this distinction in that it makes possible the claim that sentences or propositions which express ordinary semantic content can have causal relevance even if they, or the contents they express, do not have causal power, or are not taxonomized according to causal power. This move was applied to CEMS in Chapter II, and precisely the same move can be made in the case of FEMS also. For example, consider again Fodor's introduction of Weak RTM, which, in conforming to the syntactic requirement, is an expression of FEMS:

" This is the theory:

(a) Propositional attitude states are relational.

(b) Among the relata are mental representations ...

(c) Mental representations are symbols; they have both formal and semantic properties.

(d) Mental representations *have their causal roles in virtue of their formal properties.*'

Fodor's use of 'in virtue of' hides the distinction between causal relevance and causal power, hence, hides the distinction between OMS and EMS. Thus it would be consistent with (d) above to hold that the formal properties of mental representations are bearers of causal power but not

1. Representations, op. cit., p. 26.

of causal relevance, and, hence, that although mental representations have their causal roles 'in virtue of' their formal properties, those formal properties play no explanatory role in the explanations of cognitive psychology. The *Occam's Razor* response to this approach described in Chapter I is then blocked by the claim, argued for in Chapter V, that formal or syntactic properties cannot be individuated independently of ordinary semantic properties.

The positive suggestion advanced in this thesis is this: A property can have causal relevance, can play an explanatory role in causal explanations (relative to some explanatory domain) even if does not possess causal power, or is not taxonomized according to causal power, as long as it bears some appropriate relation to a property which does have causal power or is taxonomized according to causal power. This has been expressed as the claim that the causally relevant property must *index* a causally powerful property. This positive suggestion plays a minor role in the thesis, and the details are no doubt incomplete. I will conclude by mentioning a few problems which will probably have to be addressed in any future development of this view.

The first problem concerns the *conventionalism* which attends certain examples of the indexing relation. In the case of properties of number, the connection between numbers and intrinsic properties of objects is conventional. Thus, in the example given in Chapter I, the explanatory role of the number 20, where that number is attached to the kilogramme scale, could be taken over by the number 44, where that number is attached to the pounds scale. This conventionalism is

certainly not mirrored in the relation between propositions or content sentences and internal states of the brain. It is not plausible to claim that there exists only a conventional relation between our ascriptions of mental states and those states themselves. Cases such as this probably indicate that the model provided by the notion of indexing is, in itself, incomplete, and that further constraints are needed to rule out the conventionality of the indexing relation.

Secondly, as the remarks of Chapter II probably make clear, the notion of indexing, suitably developed, can and should be divorced from the notion of *isomorphism*. To begin with, Twin Earth examples make it clear that there can be no general isomorphism between, on the one hand, logical relations holding between propositions and, on the other, functional relations holding between internal states of the head. This, I believe, is sufficient to entail rejection of the condition of isomorphism in any recognizable form. Fortunately, there is nothing in the notion of indexing which entails nailing one's flag to the condition of isomorphism. It is a fact importantly in line with the central principles elicited in this thesis that for a property to have causal relevance there need be only *some* relation between that property and a causally powerful property, the relation need not be as *neat* as that entailed by isomorphism. The arguments of Chapters III and IV, and appendices II and III were intended to make this clear.

BIBLIOGRAPHY

- Armstrong, D. (1983): What is a Law of Nature? (Oxford, Oxford University Press).
- Bach, K. (1982): "De Re Belief and Methodological Solipsism", in A. Woodfield, ed., Thought and Object, (Oxford, Oxford University Press), pp. 121-151.
- Block, N. (1986): "Advertisement For a Semantics For Psychology", in Midwest Studies in Philosophy, vol. X, pp. 615-78.
- Blackburn, S. (1984): Spreading The Word. (Oxford, Oxford University Press).
- Blackburn, S. (1985): "Supervenience Revisited", in I. Hacking, ed., Exercises in Analysis. (Cambridge, Cambridge University Press).
- Blackburn, S. (Unpublished): "Losing Your Mind: Physics, Identity, and Elimination."
- Burge, T. (1979a): "Individualism and the Mental", Midwest Studies in Philosophy vol.IV: Studies in Metaphysics ed., P.A. French, T.E.Uehling, and H.K. Wettstein. (Minneapolis, University of Minnesota Press), pp. 73-121.
- Burge, T. (1982a): "Other Bodies", in A. Woodfield, ed., Thought and Object, (Oxford, Oxford University Press), pp.97-120.
- Burge, T. (1982b): "Two Thought Experiments Reviewed", Notre Dame Journal of Formal Logic, xxii, pp. 284-93.
- Burge, T. (1986a): "Individualism and Psychology", The Philosophical Review Vol. XCV, pp. 3-45.
- Burge, T. (1986b): "Cartesian Error and the Objectivity of Perception", in P. Pettit & J. McDowell eds., Subject, Thought and Context, (Oxford, Oxford University Press), pp. 117-36.
- Churchland, P.M. (1981): "Eliminative Materialism and Propositional Attitudes", The Journal Of Philosophy, 78, pp. 67-90.
- Churchland, P.S. (1986): Neurophilosophy, (Massachusetts, MIT Press).
- Davidson, D. (1963): "Actions, Reasons, and Causes", The Journal of Philosophy, vol.60, pp. 685-700. Reprinted in Davidson (1980), pp. 3-19.
- Davidson, D. (1967a): "Causal Relations", The Journal of Philosophy, 64, pp.691-703. Reprinted in Davidson (1980) pp.149-62.

- Davidson, D. (1967b): "Truth and Meaning", Synthese 17, pp. 304-23. Reprinted in Davidson (1984), pp. 17-36.
- Davidson, D. (1969): "The Individuation of Events", in N. Rescher ed., Essays in Honour of Hempel, ed., (Reidel, Dordrecht), pp. 216-34. Reprinted in Davidson (1980) pp. 163-180.
- Davidson, D. (1970): "Mental Events", in L. Foster & J.W. Swanson eds., Experience and Theory, (University of Massachusetts Press, and Duckworth, London), pp.79-101. Reprinted in Davidson (1980) pp.207-25.
- Davidson, D. (1973a): "The Material Mind", in P. Suppes ed., Logic, Methodology and the Philosophy of Science vol.4, (North Holland, Amsterdam). Reprinted in Davidson (1980) pp. 245-59.
- Davidson, D. (1973b): "Radical Interpretation", in Dialectica 27, pp. 313-28. Reprinted in Davidson (1984) pp. 125-39.
- Davidson, D. (1974a): "Psychology as Philosophy", in S.C. Brown ed., Philosophy of Psychology, (MacMillan Press, London). Reprinted in Davidson (1980) pp. 229-39.
- Davidson, D. (1974b): "Belief and the Basis of Meaning". Synthese, 27, pp. 309-23. Reprinted in Davidson (1984), pp. 141-54.
- Davidson, D. (1980): Essays on Actions and Events, (Oxford, Oxford University Press).
- Davidson, D. (1984): Inquiries into Truth and Interpretation. (Oxford, Oxford University Press).
- Davidson, D. (1986): "Empirical Content" in LePore, E. ed. Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson, (Oxford, Basil Blackwell, 1986).
- Dennett, D. (1978): Brainstorms, (Massachusetts, MIT Press).
- Dennett, D. (1980): "Reply to Professor Stich". Philosophical Books XXI, pp. 73-76.
- Dennett, D. (1981): "Three Kinds of Intentional Psychology", in R. Healey ed., Reduction, Time and Reality, (Cambridge, Cambridge University Press), pp. 37-61.
- Dennett, D. (1982): "Beyond Belief", in A. Woodfield ed., Thought and Object, (Oxford, Oxford University Press), pp. 1-95.
- Dennett, D. (1983): "Intentional Systems in Cognitive Ethology: The Panglossian Paradigm Defended". The Behavioural and Brain Sciences Vol. 6, pp. 343-90.
- Evans, G. (1980): "Understanding Demonstratives". In Parrett, H. & Bouveresse, J., Meaning and Understanding, (Berlin, New York, Walter De Gruyter).

Field, H. (1977): "Logic, Meaning, and Conceptual Role", The Journal of Philosophy LXXIV, no.7, pp. 379-409.

Field, H. (1978): "Mental Representation", Erkenntnis vol.13, pp. 9-61. Reprinted in N. Block, ed., Readings in the Philosophy of Psychology vol.2, (Massachusetts, Harvard University Press). pp. 78-114.

Field, H. (1980): Science Without Numbers: A Defence of Nominalism. (Oxford, Basil Blackwell).

Fodor, J. (1974): "Special Sciences", Synthese 28, pp. 97-115. Reprinted in Fodor (1981a).

Fodor, J. (1975): The Language of Thought, (New York, Thomas Y. Crowell).

Fodor, J. (1978): "Propositional Attitudes", The Monist, vol.61, no.4 pp. 501-23. Reprinted in Fodor (1981a), pp. 177-203.

Fodor, J. (1980): "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology.", The Behavioural and Brain Sciences, 3, 1. Reprinted in Fodor (1981a), pp. 225-53.

Fodor, J. (1981a): Representations, (Massachusetts, MIT Press).

Fodor, J. (1981b): "Three Cheers for Propositional Attitudes", in Fodor (1981a), pp. 100-123.

Fodor, J. (1982): "Cognitive Science and the Twin-Earth Problem", Notre Dame Journal of Formal Logic, vol. 23, pp. 98-118.

Fodor, J. (1986): "Individualism and Supervenience." Proceedings of the Aristotelian Society, Supplementary Volume LX, pp. 235-62. Reprinted in Fodor (1987), ch.2.

Fodor, J. (1987): Psychosemantics, (Massachusetts, MIT Press).

Follesdal, D. (1985): "Causation and Explanation: A Problem in Davidson's View on Action and Mind". In LePore & McLaughlin eds. (1985).

Frege, G. (1892): "On Sense and Reference". Reprinted in Translations From the Philosophical Writings of Gottlob Frege, ed. P. Geach & M. Black. (Oxford, Basil Blackwell, 1957).

Fuller, G. & Yu, P. (1986): "A Critique of Dennett". Synthese, vol. 66, no.3, pp. 453-76.

Harman, G. (1973): Thought. (Princeton, Princeton University Press).

Haugeland, J. (1982): "Weak Supervenience", American Philosophical Quarterly, 19, pp. 93-103.

Haugeland, J. (1983): "Ontological Supervenience", The Southern Journal of Philosophy, vol. XXII, Supplement, ed., T. Horgan (The Spindel Conference 1983: Supervenience), pp. 1-13.

Hellman, G. & Thompson, F. (1975): "Physicalism: Ontology, Determination, and Reduction." The Journal of Philosophy, vol.72, pp. 551-64.

Honderich, T. (1981): "Psychophysical Lawlike Connections and Their Problem", Inquiry, vol.24, no.3, pp.277-304.

Honderich, T. (1982a): "The Argument for Anomalous Monism", Analysis, vol.42, no.1, pp.59-64.

Honderich, T. (1982b): "Causes and if p, even if x, still q", Philosophy vol.57, pp. 291-317.

Honderich, T. (1982c): "Smith and the Champion of Mauve", Analysis, vol. 44, no.7, pp.86-89.

Horgan, T. (1982): "Supervenience and Microphysics". Pacific Philosophical Quarterly, 63, pp. 29-43.

Horgan, T. (1983): "Cosmic Hermeneutics". The Southern Journal of Philosophy vol. XXII, Supplement, ed. T. Horgan (The Spindel Conference 1983: Supervenience), pp. 19-38.

Hornsby, J. (1986): "Physicalist Thinking and Conceptions of Behaviour", in P. Pettit & J. McDowell eds., Subject, Thought and Context, (Oxford, Oxford University Press), pp. 95-115.

Jackson, F. & Pettit, P. (Unpublished): "Functionalism and Broad Content".

Johnston, M. (1985): "Why Having a Mind Matters", in LePore & McLaughlin eds. (1985), pp. 408-26.

Kim, J. (1966): "On the Psycho-physical Identity Theory", American Philosophical Quarterly, 3, pp. 227-35.

Kim, J. (1973): "Causation, Nomic Subsumption, and the Concept of Event", The Journal of Philosophy, 70, pp. 27-36.

Kim, J. (1976): "Events as Property Exemplifications", in M. Brand and D. Walton eds., Action Theory, (Dordrecht, Netherlands), pp. 159-77.

Kim, J. (1977): "Causation, Emphasis, and Events", Midwest Studies in Philosophy, 2, pp. 100-103.

Kim, J. (1978): "Supervenience and Nomological Incommensurables", American Philosophical Quarterly, vol.15, no.2, pp. 149-56.

Kim, J. (1979): "Causality, Identity, and Supervenience in the Mind-body Problem", Midwest Studies in Philosophy, 4, pp. 31-49.

- Kim, J. (1982): "Psychophysical Supervenience", Philosophical Studies, 41, pp. 51-70.
- Kim, J. (1983): "Supervenience and Supervenient Causation", The Southern Journal of Philosophy, vol. XXII Supplement, ed. T. Horgan, (The Spindel Conference 1983: Supervenience). pp. 45-56.
- Kim, J. (1984a): "Concepts of Supervenience", Philosophy and Phenomenological Research, vol. XLV, no. 2, pp. 153-76.
- Kim, J. (1984b): "Epiphenomenal and Supervenient Causation", Midwest Studies in Philosophy, 9, pp. 257-70.
- Kim, J. (1985): "Psychophysical Laws", in, E. LePore & B. McLaughlin eds., Actions and Events: Perspectives on the Philosophy of Donald Davidson, (Oxford, Basil Blackwell). pp. 369-86.
- Lennon, K. (1984): "Anti-Reductionist Materialism", Inquiry, vol.27, no. ?, pp.363-80.
- LePore, E. & Loewer, B. (1986): "Solipsistic Semantics". Midwest Studies in Philosophy, vol. X, pp. 595-614.
- LePore, E. & Loewer, B. (1987): "Mind Matters". Symposium given at Annual Meeting of The American Philosophical Association, Eastern Division. New York, December 1987. Published in The Journal of Philosophy, vol. LXXXIV, no. 11. pp. 630-42.
- LePore, E. & McLaughlin, B. (1985): Actions and Events: Perspectives on the Philosophy of Donald Davidson, (Oxford, Basil Blackwell).
- Lewis, D. (1983): "New Work for a Theory of Universals", Australasian Journal of Philosophy, vol. 61, no. 4, pp. 343-77.
- Lewis, H.A. (1985): "Is the Mental Supervenient on the Physical?", in, B. Vermazen & M. Hintikka eds., Essays on Davidson: Actions and Events, ??, pp. 159-72.
- Loar, B. (1981): Mind and Meaning. (Cambridge, Cambridge University Press).
- Lombard, L. (1979): "Events". Canadian Journal of Philosophy, vol.9, pp. 425-60.
- Lombard, L. (1986): Events: A Metaphysical Study. (London, Routledge & Kegan Paul).
- Macdonald, C. & Macdonald, G. (1986): "Mental Causes and the Explanation of Action". Philosophical Quarterly Vol. 36, no. 143, pp. 145-158.
- McDowell, J. (1985): "Functionalism and Anomalous Monism", in E. LePore & B. McLaughlin eds., Actions and Events: Perspectives on the Philosophy of Donald Davidson, (Oxford, Basil Blackwell. pp. 387-98).

- McDowell, J. (1986): "Singular Thought and the Extent of Inner Space", in P. Pettit and J. McDowell eds., Subject, Thought and Context, (Oxford, Oxford University Press). pp. 137-68.
- McGinn, C. (1978): "Mental States, Natural Kinds and Psychophysical Laws", Proceedings of the Aristotelian Society, Supplementary Volume, LII. pp. 195-220.
- McGinn, C. (1980): "Philosophical Materialism", Synthese, vol.44, pp. 173-206.
- McGinn, C. (1982a): "The Structure of Content", in A. Woodfield ed. Thought and object, Oxford, Oxford University Press. pp. 207-58.
- McGinn, C. (1982b): The Character of Mind. Oxford, Oxford University Press.
- McGinn, C. (Forthcoming): Mental Content. (Oxford, Basil Blackwell).
- McLaughlin, B. (1983): "Event Supervenience and Supervenient Causation". The Southern Journal of Philosophy, vol. XXII Supplement, ed., T. Horgan (The Spindel Conference 1983: Supervenience). pp. 71-92.
- McLaughlin, B. (1985): "Anomalous Monism and the Irreducibility of the Mental", in, E. LePore & B. McLaughlin eds., Actions and Events: Perspectives on the Philosophy of Donald Davidson. Oxford, Basil Blackwell. pp. 331-68.
- McLaughlin, B. & LePore, E. (1985): Actions and Events: Perspectives on the Philosophy of Donald Davidson, (Oxford, Basil Blackwell).
- Peacocke, C. (1981): Holistic Explanation. (Oxford, Oxford University Press).
- Perry, J. (1977): "Frege on Demonstratives". The Philosophical Review, vol.86, pp. 474-97.
- Perry, J. (1979): "The Problem of the Essential Indexical", Nous, vol.13, pp. 3-21.
- Pettit, P. (1986): "Broad-Minded Explanation and Psychology", in P. Pettit & J. McDowell eds., Subject, Thought and Context. (Oxford, Oxford University Press). pp. 17-58.
- Pettit, P. & Jackson, F. (Unpublished): "Functionalism and Broad Content".
- Post, J.F. (1983): "Comment on Teller". The Southern Journal of Philosophy, vol. XXII Supplement, ed., T. Horgan (The Spindel Conference 1983: Supervenience). pp. 163-67.

- Putnam, H. (1975a): "The Meaning of 'Meaning'". In, K. Gunderson ed., Language, Mind and Knowledge: Minnesota Studies in the Philosophy of Science vol. VII. Minneapolis, University of Minnesota Press. Reprinted in Putnam (1975b).
- Putnam, H. (1975b): Philosophical Papers vol. II, Mind, Language and Reality. (Cambridge, Cambridge University Press).
- Putnam, H. (1978): Meaning and the Moral Sciences. (London, Routledge & Kegan Paul).
- Quillen, K. (1985): "Propositional Attitudes and Psychological Explanation", Mind and Language, vol. II, pp. 133-157.
- Quine, W.V.O. (1962): Word and Object. (Massachusetts, MIT Press).
- Robinson, H. (1982): Matter and Sense. (Cambridge, Cambridge University Press).
- Smart, J.J.C. (1985): "Davidson's Minimal Materialism", in Vermazen & Hintikka (1985), pp. 173-82.
- Smith, P. (1982): "Bad News for Anomalous Monism?". Analysis, vol. 42, no. 7, pp. 220-24.
- Smith, P. (1984): "Anomalous Monism and Epiphenomenalism: A Reply to Honderich". Analysis, vol. 44, no. 2, pp. 83-86.
- Sosa, E. (1984): "Mind-Body Interaction and Supervenient Causation". Midwest Studies in Philosophy, 9. pp. 271-81.
- Stalnaker, R. (1984): Inquiry, (Cambridge, Massachusetts, MIT Press).
- Stanton, W. (1983): "Supervenience and Psychophysical Law in Anomalous Monism". Pacific Philosophical Quarterly, vol. 64, no. 1, pp. 72-79.
- Sterelny, K. (1984): "Critical Review of A. Woodfield ed., Thought and Object". Journal of Semantics, vol.3, pp. 277-94.
- Stich, S. (1978): "Autonomous Psychology and the Belief-Desire Thesis". The Monist, vol.61. pp. 573-91.
- Stich, S. (1982): "On the Ascription of Content". A. Woodfield ed., Thought and Object, Oxford, Oxford University Press. pp. 153-206.
- Stich, S. (1983): From Folk Psychology to Cognitive Science: The Case Against Belief. Massachusetts, MIT Press.
- Stoutland, F. (1976): "The Causation of Behaviour". Essays in Honour of G.H. Von Wright (Acton, Philosophica Fennica), pp. 286-325.
- Stoutland, F. (1985): "Davidson on Intentional Behaviour". in LePore & McLaughlin eds. (1985).

Swoyer, C. (1982): "The Nature of Natural Laws". Australasian Journal of Philosophy, vol. 60, no. 3, pp. 203-23.

Teller, P. (1983a): "A Poor Man's Guide to Supervenience and Determination". The Southern Journal of Philosophy vol. XXII Supplement, ed. T. Horgan (The Spindel Conference 1983: Supervenience). pp. 138-62.

Teller, P. (1983b): "Comments on Kim's Paper". The Southern Journal of Philosophy vol. XXII Supplement ed., T. Horgan (The Spindel Conference 1983: Supervenience). pp. 57-61.

Tooley, M. (1977): "The Nature of Natural Laws". Canadian Journal of Philosophy Vol VII, pp. 667-98.

Vermazen, B. & Hintikka, M. eds., (1985): Essays on Davidson: Actions and Events. (Oxford, Oxford University Press).

White, S. (1982): "Partial Character and the Language of Thought", Pacific Philosophical Quarterly, vol. 63, pp. 37-65.

Yu, P, & Fuller, G. (1986): "A Critique of Dennett", Synthese, vol.66, no.3, pp. 453-76.