

Mental States and Psychological Explanation

PEI Kong-ngai

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Master of Philosophy

in

Philosophy

©The Chinese University of Hong Kong

September 2002

The Chinese University of Hong Kong holds the copyright of this thesis. Any person (s) intending to use a part or whole of the materials in the thesis in a proposed publication must seek copy release from the Dean of the Graduate School.

During the past few decades, the study of individual differences in psychology has become an important part of the discipline. The study of individual differences is not only a theoretical problem, but also a practical one. The theoretical part of the present study is to examine the properties of individual differences in detail. Some philosophical questions which can be derived from certain theoretical principles, in this regard, however, I shall argue that individualism cannot be said to have a solid ground, nor does it conform to actual practices in the sciences. As a result, the above arguments for individualism are unsound.



To Lorraine

the study of individual differences in psychology has become an important part of the discipline. The study of individual differences is not only a theoretical problem, but also a practical one. The theoretical part of the present study is to examine the properties of individual differences in detail. Some philosophical questions which can be derived from certain theoretical principles, in this regard, however, I shall argue that individualism cannot be said to have a solid ground, nor does it conform to actual practices in the sciences. As a result, the above arguments for individualism are unsound.

香港中文大學圖書館藏書

YAM 30
0 8 MAY
THE CHINESE UNIVERSITY OF HONG KONG
圖書館



Abstract

During the past two decades, the ontological status of the theoretical posits of folk psychology – intentional properties – has invited a great deal of debate. According to the doctrine called “eliminativism,” intentional properties such as beliefs and desires are not real or genuine. One argument for eliminativism can be summarized as follows: (1) Intentional properties do not supervene on intrinsic physical properties of individuals. (2) However, individualism holds that the theoretical posits of scientific disciplines must be those determined by intrinsic physical properties of individuals. (3) The theoretical posits of the sciences exhaust all existents. Thus, (4) intentional properties do not exist. The central aim of this thesis is to examine premise individualism in detail. Some philosophers (e.g., Fodor) hold that individualism can be derived from certain metaphysical principles. In this thesis, however, I shall argue that individualism cannot be derived from a priori grounds, nor does it conform to actual practices in the sciences. As a result, the above argument for eliminativism is unsound.

摘要

近二十來年，大眾心理學的理论元目—意何性質—一直引起大量的爭論。根據消除性唯物論者的觀點，意何性質如信念及欲望等並不屬真實的存有物。消除性唯物論的其中一個論證可被勾劃如下：(1) 意何性質並不從屬於個體的內在物理性質。(2) 然而，個體主義宣稱，科學的理论元目必需被個體的內在物理性質決定。(3) 科學的理论元目窮盡一切真實存有物。故此，意何性質並不存在。本文的主要目的是要對個體主義作一詳細考察。一些哲學家（如福特）認為，個體主義可由某些形而上的原則導出。然而本文將指出，個體主義不僅不能被先驗地導出，此外，個體主義亦不符合科學的實際運作。故此，以上消除性唯物論的論證不能成立。

Table of Contents

Chapter One: Introduction

1. Intentional States and Folk Psychology	1
2. Eliminativism, Externalism, and Individualism	5
3. Overview of The Thesis	8

Chapter Two: Mental States and Externalism

1. The Distinction between Intrinsic and Relational Properties	11
2. Supervenience	14
3. Externalism	18
4. The Classical Arguments for Semantic Externalism: The Twin Earth Thought Experiments	19
5. From Semantic Externalism to Mental Content Externalism	23
6. Externalism and Physicalism	26
7. The Common Concept Strategy Objection to Externalism	28

Chapter Three: Stich's Argument for Individualism: The Argument from Below

1. Stich's Argument for Individualism	34
2. Narrow and Wide Behaviour	37
3. Refining the Argument	39
4. Is Replacement Argument Successful in establishing Individualism?	44
5. Fodor's Argument for Premise 2*: Narrow Behaviour and Crazy Causal Mechanisms	46
6. Causal vs. Non-Causal (Constitutive) Causal Powers	51
7. Conclusion: Stich's Unsuccessful Argument from Below	53

Chapter Four: Fodor's Argument for Individualism: The Argument from Above

1. Fodor's Explanan Argument	56
2. A Response to Individualism: Rendering Intentional States Individualistic	65
2.1 Fodor's Account of Narrow Content	69
2.2 Criticisms of Fodor's Account of Narrow Content	71
3. Examining Global Individualism: Fodor's A Priori Argument	74
3.1 Counterexamples to Global Individualism	76
3.2 Can Global Individualism be Reconciled with Relational Taxonomies?	81

3.3 Two Senses of Causal Relevance of External Conditions	83
3.4 The Failure of The Argument from Above	89

Individualism

Chapter Five: Conclusion	91
--------------------------	----

Bibliography	i
--------------	---

Acknowledgement	vi
-----------------	----

Chapter 1. Introduction: The Eliminativist's Argument from Individualism

1. Intentional States And Folk Psychology

In daily life, we often ascribe other people mental states such as beliefs and desires to explain and predict their behaviour. Here is a typical example. Suppose that we saw Lorraine open the fridge and are interested in why she did so. To explain her action, we can rationalize or make sense of it in a way like this: She opened the fridge because she desired to have a glass of water and believed that opening the fridge would help her satisfy her desire. If, to modify the example, we instead know in advance what Lorraine wants and believes, we can predict that she will tend to open the fridge.

States such as beliefs and desires (along with intentions, hopes, wants, and other attitudes) belong to a category of mental states often called "*intentional states*".¹ Generally, intentional states are conceived as involving an individual's attitude toward a proposition or truth condition (this is why they sometimes are called "propositional attitudes"). The *attitudes* include such things as belief, desire, fear, hope, and so on. The proposition in question constitutes the state's content. In general, the form of

¹ Some authors distinguish the terms "intentional properties" and "intentional states", using the former to refer to instantiations of the latter at particular times. In what follows, when there is no threat of confusion, I shall use the two interchangeably.

statement that ascribes a propositional attitude to an individual can be characterized as follows: “SΨ-s that p”, where “S” refers to an individual, “Ψ” names any verb expressing an attitude, and “that p” specifies any proposition or truth-condition. “Lorraine believes that her boyfriend is smart” and “Pei hopes that Lorraine believes that her boyfriend is smart” are two typical examples of such statements.

It is generally believed that intentional states have the following characteristics.² First, intentional states have contents or meanings, which are semantically evaluated with respect to the world – that is, they can be true or false. For example, my belief that the sun will rise tomorrow has *the sun will rise tomorrow* as its content, and is a true belief. Typically the content of an intentional state is specified by the *that clause*, in this case, “that the sun will rise tomorrow”.

Second, intentional states are *individuated by their attitudes and contents*. Hoping that today is a rainy day and believing that today is a rainy day, though having the same content – that *today is a rainy day* – count as two different intentional states, since they are different attitudes. The belief with the content that *I will pass the exam* and the one with the content that *I will not pass the exam* are also regarded as two different intentional states, for though they are the same kind of attitude, they have different contents.

² See, e.g., Fodor (1987). ch. 1.

Third, intentional states are causally efficacious³ with respect to behaviour and other mental states, and so are regarded as having *genuine causal powers* – that is, having effects on the physical world. We expect that changes in mental states will produce certain changes in behaviour and other mental states. To illustrate, consider again the above case. Intuitively, it is because Lorraine has intentional states with the contents they do that she exhibited a particular behaviour (and other mental states). Had she not possessed intentional states with those contents – that is, not possessed any (occurrent) intentional states at that time, or possessed mental states with other contents – she would not have behaved in that way.

According to a view often called “Theory theory”, we are capable of predicting and explaining people’s behaviour in terms of intentional states because we apply an implicit empirical theory sometimes called “folk psychological theory (or just “folk psychology”).⁴ In general terms, we may think of a theory as a set of principles or generalizations for explaining certain phenomena. Newtonian mechanics, for example, can be thought of as a cluster of principles devised to explain the motions of physical objects. That folk psychology is a theory, thus, implies that it comprises a collection

³ For variety, in what follows I shall use “causally irrelevant (to behaviour and other mental states),” “causally inefficacious,” “causally impotent,” “causally inert,” and “causally redundant” interchangeably.

⁴ In the current literature, “folk psychology” is just one name among many to name this implicit theory. Some prefer the name “commonsense psychology.” Other labels include “belief-desire psychology,” “naïve psychology.” However, since folk psychology is still the most commonly used label in the literature, in what follows I shall conform to this practice.

of psychological generalizations aiming at explaining mental phenomena and behaviour. Where *folk* psychology is concerned, these psychological generalizations are so platitudinous and intuitive that “everyone knows them, everyone knows that everyone else knows them, and so on” (Lewis (1972): p.256). Some such generalizations are as follows:

- a. When a normal person is looking at a traffic light, which changes from red to green, she usually comes to believe that it has changed from red to green.
- b. If a person believes that all scorpions are poisonous, and if she comes to believe that Henry’s pet is a scorpion, then she will typically come to believe that Henry’s pet is poisonous.
- c. If a person sitting at a bar wants to order a beer, and if she has no stronger desire to do something that is incompatible with ordering a beer, then typically she will order a beer.⁵
- d. If a person desires to eat a cake and believes that opening the fridge can help her realize her desire, she will tend to open her fridge.

On this view, we can successfully explain Lorraine’s behaviour because the event to be explained, her opening the fridge, is subsumed under d. Note although that the

⁵ a – c are from Stich and Ravenscroft ((1995); Stich (1996). p.126)

psychological laws given above refer to particular contents of mental states (e.g., *the desire to order a beer*), they are not *content-dependent*.⁶ Rather they can all be regarded as instances of more general laws. For instance, (d) can be regarded as an instance of a general law we might call “the practically reasoning syllogism”:
(x)(D)(Q) (If x desires that D, and believes that by doing Q, x can succeed in realizing D, then, *ceteris paribus*, x will do Q).

2. *Eliminativism, Externalism, and Individualism*

During the past two decades or so, the ontological status of theoretical posits of folk psychology – intentional states – has invited a great deal of debate. According to the doctrine called “eliminative materialism” (or just “*eliminativism*”), just as “witches and Homer’s gods are just myths” (Stich (1995); (1996): p.115), beliefs and desires do not exist.⁷ At least at first sight, no doctrine in philosophy of mind is as radical and shocking as eliminativism.⁸ In explaining and predicting behaviour by making use of intentional states, we seem to presuppose that some explanations of this kind are true, and some are false. But if eliminativism is right, it follows that none of these

⁶ But one should note that not all laws of folk psychology are of this kind. Some psychological laws inevitably involve particular contents or types of contents. The law that the moon looks bigger near the horizon, and that people will act in order to avoid a perceived threat are two instances of content-dependent laws (see Botterill and Carruthers (1999)).

⁷ One should note that not all authors use the term “eliminativism” to refer to this ontological claim about intentional states; some use it just for the view that folk psychology is a radically mistaken theory. For the relationship between these two readings of “eliminativism,” see (Stich (1996)).

⁸ The following alleged implications of eliminativism are pointed out by Stich and Ravenscroft (1995).

explanations can be true, for there are simply no such things as intentional states.

Moreover, many disciplines presuppose the existence of intentional states. For example, for a long time whether knowledge is justified true belief has been one of central problems in epistemology. But now again suppose that eliminativism turns out to be true, then all our attempts to answer this problem become worthless, since there is no such thing as a belief. When reflecting deeply on the implications of eliminativism, we may tend to agree with Jerry Fodor that the success of eliminativism would be “the greatest intellectual catastrophe in the history of our species” (1987: xii).

There are various routes to eliminativism. In this thesis I shall examine one often discussed argument for eliminativism: *the argument from individualism*.⁹ It goes roughly as follows: According to the doctrine often called *externalism*, intentional properties are widely individuated, which, means roughly that their identity conditions depend not just on *intrinsic* or *narrow* properties but also on certain *relational* or *wide* properties. Scientific psychology, the subject that aims to tell us the true story about mind, must abide by *individualism*, which holds roughly that the explanatory categories or kinds taxonomized or individuated by scientific psychology must be narrow. As a result, at least at first sight, intentional states as folk psychology

⁹ See Stich (1996): pp.22-3, Botterill and Carruthers (1999): pp.155-6.

conceives them cannot have any legitimate role in scientific psychology. This claim, accompanied by the controversial but common assumption that *only* entities and properties posited by the sciences are real or genuine, some philosophers argue, implies eliminativism. The argument can be stated more clearly as follows:

The Eliminativist's argument from Individualism

1. Folk psychology is an implicit theory that predicts and accounts for how the mind works, i.e., how certain mental states produce other mental states and how interactions between mental states produce behaviour. It consists of a set of generalizations couched in terms of intentional properties.
2. Intentional properties, the theoretical posits of folk psychology, do not supervene on intrinsic physical properties.
3. The theoretical posits of scientific psychology, however, must be properties that supervene on intrinsic physical properties.
4. Therefore, the intentional states of folk psychology cannot be proper candidates for theoretical posits of scientific psychology (from 1, 2, and 3).
5. Science is the only tribunal of what things are real or what exists.
6. Therefore, intentional states such as beliefs and desires do not exist. Or, equivalently, intentional properties are not real or genuine (from 4 and 5).

To be sure, all the premises of the above argument (1, 2, 3, and 5) are controversial and the argument establishes the eliminativist's conclusion only if all its premises turn out to be true. *In this thesis, I will not attempt to evaluate all the premises, but will focus mainly on premise 3.*

3. Overview of the thesis

The next chapter is devoted mainly to explaining the meaning of premise 2, that intentional properties do not supervene on intrinsic physical properties. I will explain several notions, such as intrinsic and relational properties, which are essential to understanding externalism properly. Externalism can be derived in various ways. The most well-known of these is the Twin-Earth thought experiments presented by Putnam in his celebrated paper "The meaning of "meaning"" (1975). He argues that ascriptions of intentional states to an individual need to take account of the individual's environment. Though my major concern is not whether externalism is true, in this chapter I shall examine one famous objection to it, since doing so may help clarify the nature of this doctrine.

Chapters 3 and 4 are the core of the thesis. In chapter 3, Stich's often-discussed argument for individualism, what he calls "the replacement argument", will be examined in detail. Since it derives individualism mainly from a normative claim

concerning the explananda of psychology – that scientific or systematic psychology should concern itself *only* with explaining and predicting autonomous behaviour – I call it “the aim of psychology argument”. In that chapter I shall argue that this normative doctrine does not do justice to actual practices of psychological inquiry. Since Stich provides us with no *a priori* grounds for it, I conclude that his argument fails.

However, I think that an argument in Fodor’s *Psychosemantics* (1987) concerning the nature of causation can be used to support this normative doctrine. This argument is a *reductio* one. It argues that taxonomizing the explananda of psychology non-autonomously (indeed the explananda of any science) needs to postulate a kind of causal mechanisms. Since Fodor thinks that such mechanisms are mad or impossible, the explananda of psychology must be taxonomized autonomously. However, I argue that Fodor’s argument does not work, for he confuses causation with individuation.

In chapter 4, I shall discuss Fodor’s argument, which I call “explanan argument”. His argument is so called by me because it argues for individualism not from a normative constraint on the individuation of psychological explananda but from a claim about how explanantia should be typed. Fodor argues that any taxonomy of explanatory kinds (explanantia) in the sciences must obey a constraint that he calls “global individualism”, which holds that explanatory kinds devised by the sciences

must be individuated by their causal powers: If two things agree in causal powers, then they must be grouped under the same kind. Since psychology is generally regarded as a scientific discipline that provides causal explanations and predictions by constructing causal generalizations, it must as well obey global individualism. The causal powers of anything, according to Fodor, are determined by or supervenient on that thing's intrinsic physical properties. It follows that the causal powers of intentional states, the explanantia of scientific psychology, must supervene on intrinsic physical properties.

Though Fodor thinks that explanatory kinds in psychology must be individualistic, as a friend of folk psychology, he tries to remove the conflict between premise 2 and 3 by working out a notion of content (*narrow content*) from content employed in folk psychology. I argue that Fodor's attempt is unsuccessful.

Fodor thinks that global individualism is supported not only by actual practices of individuation in scientific disciplines but also by an uncontroversial view about causation. I argue that, to the contrary, global individualism cannot do justice to actual scientific practices. Nor can it be derived from metaphysical grounds. As a result, it fails to place a plausible constraint on taxonomy of explanantia. Worse, I argue that Fodor's explanan argument cannot be sound, since it either contains false premises or equivocates on the notion of causal powers.

Chapter 2: Mental States and Externalism

This chapter is devoted to explaining the meaning of externalism. In section 1, I shall explain the distinction between intrinsic and relational properties. Understanding it is essential to properly understanding externalism. Since the notion of supervenience can help characterize the idea of the intrinsic and the relational, in section 2, a brief account of supervenience will be provided. In sections 3 to 5, I will present the standard argument for externalism – the Twin-Earth thought experiment and explain how externalism with regard to intentional states can be derived from semantic externalism. In section 6, one alleged threat of physicalism will be discussed. In section 7, I shall discuss one common objection to externalism, which arises from misunderstanding the nature of externalism.

1. The Distinction between Intrinsic and Relational Properties

Some properties are *intrinsic*.¹⁰ Crudely, a property counts as an intrinsic one just in case the possession of it by an object is *wholly determined by the object itself*.

Paradigmatic narrow properties include such things as physiological, chemical, and

¹⁰ A terminological note: in the current literature various names are used to denote the intrinsic and relational properties characterized here (e.g., narrow - wide; autonomous - non-autonomous; internal – external (for variety, sometimes I will also use these terms)). Moreover, philosophers usually use the term “intrinsic physical properties” to denote a kind of intrinsic property – those posited by fundamental physics (e.g. Wilson (1995)) and call any properties that supervene on this kind of property “autonomous properties” (e.g. Stich (1978), (1983)).

microstructural constitutions. The particular molecular structure of my ring, for instance, is an example of an intrinsic property, for whether or not an object possesses this property is a matter of how it is, wholly independently of things external to it¹¹. Similarly, my book shelves possess the property of being composed of a certain chemical composition, which is intrinsic to the shelves and involves nothing external to them.

I shall call properties that are not intrinsic "*relational*". As a first approximation, a property can be said to be relational just in case *the possession of it is determined not only by the object itself but also by things external to it*. As a result, it is possible that there are two objects indiscernible with respect to their intrinsic features, yet only one of them possesses a certain relational property. The properties of being located ten kilometers from Fung King Hey building, being embedded in a world in which there is H₂O in rivers, and living in a speech community in which the speakers use "arthritis" to refer to the inflammation of ankle, are typical examples of relational properties, since whether or not an object possesses them depends on *the natural and social environment* it is in.

Some properties are relational because an object's possessing them partly depends on *historical* factors. One obvious example is being a dime. An object can be a dime

¹¹ In saying this I am not denying that my ring can causally interact with the external world and hence its molecular structure can change over time.

only if it possesses certain historical properties, namely being certified by the government. Anything that doesn't possess such properties is not a dime, but at most a fake. Another example is being a mosquito bite. A mark on the skin created by microsurgery is not a mosquito bite, even if it is intrinsically indiscernible from a real one.

Some relational properties have their own *narrow counterpart or correlate*. That is, they are *hybrid* relational properties in the sense that they can be decomposed into two components, one an intrinsic property, the other a relational one. One obvious example of such properties is weight. Weighing two newtons is a relational property. Yet corresponding to this property, there is an intrinsic property, having a mass of a certain number of kilograms, which is correlated with the relational property of being subject to a gravitational field with certain values, in such a way that these properties jointly constitute the property of weighing two newtons. Another uncontroversial example of a hybrid relational property is being a planet. Something is a planet in virtue of facts about its possession of a certain physical constitution (the internal or intrinsic component of being a planet) *and* facts about its orbiting a particular star (the external or relational component of being a planet). Anything that does not possess both of these components cannot be a planet. By contrast, to possess a *pure* relational property, such as living in a watery world, an object need *not* possess any particular

physical composition. Whether an object possesses a pure relational property depends *wholly* on whether it possesses certain environmental factors.

The above characterizations of narrow and wide properties are somewhat rough.

To make them more precise, we can make use of the notion of supervenience. Let me provide a brief account of supervenience.

2. *Supervenience*

Supervenience is intended to capture the idea that one set of properties is *wholly determined by* or *dependent on another*.¹² The idea of supervenience can be formulated in various ways. One common formulation is by means of the notion of *indiscernibility*. To illustrate, let **S** and **B** be two sets of properties (call them “supervenient properties” and “base properties”, respectively). Then **S** *weakly supervenes* on **B** just in case:

It is *necessary* that for any two objects **a**, and **b**, if they are indiscernible with respect to properties in **B**, then they will be indiscernible with respect to all properties in **S**.

¹² Other than capturing the notion of dependency or determination, supervenience is usually expected to explicate the notion of *covariance* and to do justice to the *nonreducibility* of the supervenient properties to their base properties. Whether or not supervenience can do all of these jobs is still a controversial question. For a more detailed discussion of supervenience see Kim (1990).

The notion of “necessity” is usually elucidated by the language of possible worlds.¹³

Crudely, to say that a statement is necessarily true is to say that it is true in all possible worlds. The above formulation thus is just to say that whatever possible world that we pick out, if we find that **a** and **b** instantiate all properties in set **B**, we will also find that they will not differ in any property in the set **S**.

Weight, for example, supervenes on mass and local gravity. For in all possible worlds, no pair of objects having the same mass and subject to the same local gravity have different weights. By contrast, weight does not supervene on size, since in some possible worlds, some objects have the same size but differ in weight.

It is generally agreed that supervenience can also be formulated by employing *nomological relations*. With certain mereological assumptions, the following formulation is generally regarded as logically equivalent to the one in terms of the notion of indiscernibility.¹⁴

S weakly supervenes on B just in case it is necessary that for any property **F** in **S**, if an object **a** has **F**, then there is a property **G** in **B** such that **a** has **G** and if any other

¹³ Since the nature of possible worlds can be specified in a variety of ways (*logical, conceptual, or nomological possible world*), forms of supervenience vary depending on what notion of possible worlds we choose.

¹⁴ See, e.g., Kim (1996), (1998).

object has **G**, it has **F**.

As I said above, weight supervenes on mass and local gravity. So for any possible world we select, if an object **a** has the weight of 10 newtons, then it has a certain mass and is subject to local gravity with a certain value such that if any other object *in the same possible world* instantiates those properties, it weighs 10 newtons as well.

One should note that *weak supervenience is not valid across possible worlds, but just relative to a particular one*. As a result, it permits two objects in different possible worlds to differ in their supervenient properties even if they share all the same base properties. Or, to put it differently, it is compatible with weak supervenience that a given property in **S**, **F**, has a property **G** as its supervenient base in a possible world w_1 , and yet there is another possible world, w_2 , in which an object, though instantiating **G**, does not possess **F**. For example, that weight only weakly supervenes on mass and local gravity is compatible with the claim that two objects in two different possible worlds, which have the same mass and are subject to the same local gravity, differ in their weight (the object in w_1 weighs x , but the one in w_2 weighs y), or that a given weight x has a certain set {mass x^* , local gravity z } as its supervenient base in w_1 , but some object in w_2 does not weigh x even though it instantiates the same mass-gravity set.

We can construct a stronger version of supervenience if the validity of supervenience holds relative not to a particular possible world, but across all possible worlds. Supervenience so construed is called “strong supervenience.” For this notion, the indiscernibility formulation is as follows:

S strongly supervenes on B just in case for any world w_1 and w_2 , and for any objects **a** and **b**, if **a** in w_1 and **b** in w_2 are indiscernible with respect to all **B** properties, then **a** in w_1 and **b** in w_2 are indiscernible with respect to all **S** properties.

And we obtain the nomological relation formulation of strong supervenience by adding a second “necessarily” to the weak one:

S strongly supervenes on B iff it is necessary that for any property **F** in **S**, if an object **a** has **F**, then there is a property **G** in **B** such that **a** has **G** and *necessarily* if any other object has **G**, it has **F**.

So much for the notion of supervenience. With the help of this notion, the notion of a property being determined by the others can be made more precise and thus the

notion of narrow and wide properties can be made more precise, as follows. A property **P** is said to be narrow iff it supervenes on intrinsic physical properties. By contrast, a property **P** is said to be wide iff it does not supervene only on intrinsic physical properties.

3. Externalism

There are debates concerning whether a given kind of properties is intrinsic or relational. I shall call the doctrine that a given kind of properties **K** is intrinsic “individualism (with regard to **K**)” and the doctrine that a given kind of properties **K** is relational “externalism (with regard to **K**)”.

There should be no doubt that externalism with regard to certain kinds of intentional states must be true. Remembering that **p** (**p** stands for any proposition), for example, is a relational intentional state. Whether or not you remember something is in part dependent on your history, in that to remember an incident, you must exist at the time when that incident occurred. Another example of relational intentional states is knowing that **p**. For, Gettier problem aside, a justified belief about a certain state of affairs counts as knowledge only if it is true, and whether or not a belief is true, of course, hinges on the environment that the epistemic subject is in.

What is controversial is whether externalism can be extended to intentional states

such as beliefs, desires, and intentions. As I said in section 1, some writers hold the view that externalism about beliefs and desires is also true. To argue for this doctrine, externalists regarding intentional states typically take two steps. First, they argue that the meanings of certain kinds of words are widely individuated (we can call this thesis “semantic externalism”). Then, with certain plausible assumptions, they argue further that beliefs involving the concepts corresponding to those words are also wide properties. The most celebrated argument for the truth of semantic externalism is the Twin-Earth thought experiment invented by Hilary Putnam (1975).¹⁵

4. The Classical Argument for Semantic Externalism: The Twin-Earth Thought Experiment

Imagine that we go back to a time prior to modern chemistry and that there is a planet, Twin-Earth, elsewhere in the universe that is exactly like the earth we live on except for one feature. On Twin-Earth there is no H₂O, but there exists a chemical compound XYZ, the observable and functional properties of which (e.g., being colourless, able to quench thirst, and tasteless), are indistinguishable from the stuff we on Earth call “water.” Moreover, suppose that each of the inhabitants of Earth has a twin or duplicate on Twin-Earth, who is type-identical to each earthling down to the last

¹⁵ There are some different forms of externalism not established by twin-earth thought experiments (e.g., G. Evan (1982), J. McDowell ((1984), (1986)). Whether or not they are true is an interesting question. However, I will not discuss them in this thesis.

elementary particle,¹⁶ and that these Twin-Earth doubles call XYZ “water.”

Now suppose that an earthling, Oscar, utters the sentence “Water is a boring drink,” and his twin, Oscar*, on Twin-Earth utters “Water is a boring drink.” Do their utterances mean the same thing? Putnam thinks not, for scientists tell us that water is H₂O, and since XYZ is not H₂O, it is not water. Consequently, when Oscar and his twin use the word “water”, they are referring to different substances. In other words, the reference or extension of the word “water” in our mouths and in twins mouth is different: For earthlings “water” refers to H₂O, whereas for twin-earthling, it refers to XYZ. And thus we should dub the compound on twin earth with a new name, say “twater”. Now since the meaning of an utterance is partly determined by the words involved in it, it follows that Oscar’s utterance and her twin’s have different meanings.

One of the main aims of the twin earth thought experiment is to show the incoherence of two theses endorsed by the traditional notion of meaning. The first thesis is “that knowing the meaning of a term is just a matter of being in a certain psychological state” (Putnam (1973); (1993): p.151). That is, the meanings of words are *entirely fixed* by the *concepts* or *intentions* with which the speakers associate them. The second is the doctrine that *meaning determines reference*: that two terms have

¹⁶ Following many philosophers, I ignore the fact that whereas twins have XYZ in their bodies, we have H₂O.

different references, that is, denote different things, implies a difference in their meanings.

If we agree with Putnam's argument in the Twin-Earth story, we then have a case in which two speakers share all the same intrinsic physical properties and hence have the same narrow psychological states, yet the meanings of their utterances differ. Thus, if meaning, as the traditional theory suggests, is construed as a concept, which is a narrow psychological state, then it does not determine reference. This means that the two assumptions on which the traditional view of meaning rests cannot both be true. Consequently, if we want to preserve the thesis that meaning determines reference, as Putnam does, the only way to do so is to enlarge the meaning of the "meaning" of words in such a way that it includes certain external factors, in this case, the environmental factors of living in a watery world. But if so, the meanings of natural kind terms are "*world involving*,"¹⁷ in that what one means by terms such as "water" depends on things external to one's body.¹⁸ In summary, *meaning is not only in the head*, as Putnam puts it.

Note that Twin-Earth experiment concerns only the question whether the meaning of one kind of term, those referring to things where a distinction can be made

¹⁷ The term "world involving" is borrowed from Woodfield (1982). I take this word to be synonymous with "widely individuated."

¹⁸ Note that since twin earth and our earth are generally conceived to be in the same possible world or universe, the above twin earth thought experiment just establishes the claim that intentional properties do not weakly supervene on intrinsic physical properties.

between the thing's appearance and the microstructure that explains its appearance, are widely individuated. I shall call this kind of term "natural kind terms" (they are sometimes also called "substance words"), and externalism concerning the meaning of such terms "*natural kind term externalism.*"

Some philosophers are quick to point out that with a little trick, we can employ Twin-Earth style thought experiments to show that meanings of words other than natural kind terms are also widely individuated. One such attempt is provided by Tyler Burge (1979).¹⁹ It goes roughly as follows:

Suppose that because of some sort of confusion or misinformation, Alf, an inhabitant of our earth, erroneously believes that "arthritis" means inflammation not only of joints but also of bones. Feeling pain in his thigh, he asserts, "I have arthritis in my thigh." Since in our linguistic community, "arthritis" cannot apply to bones but only to joints, his utterance is false. Now suppose Alf*, the physical replica of Alf who lives in a twin earth, where all things are pretty much the same as on earth except that the linguistic community there uses "arthritis" to mean inflammation of joints and bones, produces the same sort of utterance. His utterance should be regarded as true. But if Alf's utterance is false, whereas Alf*'s true, then we should conclude that the

¹⁹ Note that Burge does not establish externalism with regard to terms by using twin earth experiment. His original formulation, rather, is to consider how the arthritis sufferer's utterance and belief should be ascribed in the *actual* situation, and in a *counterfactual* situation in which the language convention of the usage of arthritis is different from the actual one.

contents or truth conditions of their utterances are different. Since *ex hypothesis*, Alf and his replica agree in their intrinsic physical properties, we now have a case in which two individuals are identical with regard to intrinsic physical states, yet their utterances, though of the same type, mean different things. And since in this case the natural environments that the physical twins live in are relevantly identical, the only difference being their linguistic community, we have shown that meanings of words are determined not only by the natural environment but also by the linguistic community or social environment. I shall call this kind of externalism “social externalism”.

5. From Semantic Externalism to Mental Content Externalism

Thus far we have an externalist account only of the semantics of terms. Some philosophers, however, take the moral of the Twin-Earth experiment one step further and use it to establish externalism with regard to the intentional properties posited by folk psychology.²⁰ They argue that intentional properties are individuated by contents, and contents, according to folk psychology, are construed as truth conditions or meanings, typically specified by means of words. Thus, assuming that when a speaker is sincere, what she means by an utterance is the content of what she believes,²¹ the

²⁰ See, e.g., C. McGinn (1977), T. Burge (1979).

²¹ Some philosophers deny this inference (e.g., R. Cummins (1991)).

externalist's conclusion about meaning can thus be carried over to beliefs (and perhaps to other intentional states such as desires and intentions). As a result, the twin earth thought experiments establish not only that meaning is "world involving," but also that intentional states are partly individuated by the environment that the individual is in.

For clarity, let's formulate the twin earth arguments for externalism with regard to intentional states in the following way:

1. In the twin-earth story, when using the word "N", the physical twins (e.g., Oscar-Oscar* and Alf-Alf*) refer to different things and so their utterances involving "N" have different truth conditions or meanings.
2. When a speaker is sincere, the meaning of her utterance is the same as the content of her belief.
3. Intentional states such as beliefs are individuated by their contents.
4. The identity condition of a belief (type) is determined by or supervenient on the content of the belief. That is, if two belief tokens are of different types, then they have different contents or meanings.
5. Thus, the physical twins have different beliefs.
6. The physical twins, ex hypothesis, are indiscernible with regard to their intrinsic

physical states.

7. Thus, the physical twins, even if they are intrinsically physically exactly alike, have different beliefs.
8. Therefore in some cases two individuals may share all the same intrinsic physical properties but not all the same intentional states.

More precisely, we can formulate externalism²² in terms of the following supervenience thesis:

It is *not necessary* that, for any two individuals x and y , and any intentional property M , if x and y are indiscernible with regard to all of their intrinsic physical properties, then x and y are indiscernible with regard to M .

But if the above supervenience thesis fails, then it follows that the belief that an individual instantiates is partly constituted by the environment in which she is embedded. In short, intentional properties are widely individuated.

²² A terminological note: the term “externalism” is ambiguous. Sometimes it is used by some authors to refer to the doctrine that intentional states *are* widely individuated according to our folk psychological practices. Yet at times it is used to denote the normative thesis – (sometimes) intentional states *should* be widely individuated *for the purposes of systematic psychological explanations*. Some philosophers (e.g. Fodor (1987); (1991)) agrees that externalism construed as the actual claim is true, yet denies that it as a plausible normative claim. To avoid confusion, in what follows I shall preserve the name “externalism” for the doctrine about how intentional states are actually individuated and call the normative doctrine “anti-individualism.”

Some philosophers remain unconvinced that the twin earth thought experiments successfully establish the doctrine of externalism with regard to intentional properties. Nevertheless, I do not intend to get embroiled in these issues, since as I said in the Introduction, the aim of this thesis is to examine whether individualism is true in psychology. Moreover, the individualists (Stich and Fodor)²³ that I will discuss in the next two chapters take the above argument as decisive. In section 7, however, I shall examine one objection (often called “the common concept strategy”) to the above argument, for it can help us to properly understand the nature of the doctrine of externalism. Before doing this, I shall first deal with another objection to externalism, which arises from a misunderstanding of physicalism.

6. Externalism and Physicalism

At first sight, there appears to be a conflict between externalism and physicalism about the mental. Very roughly, (ontological) physicalism is the view that mental properties must in some sense hinge on physical properties (those posited by the best current theories of fundamental physics). Having rejected central state theory or the type identity theory of mind,²⁴ some philosophers think that physicalism should be

²³ See, e.g., Stich (1978) and Fodor (1987: ch 2).

²⁴ Crudely, the type identity theory can be formulated as follows. For any mental property M (pain), if an object o instantiates it, then there exists a brain property B (e.g., c-fiber activation) such that o instantiates M at time t iff it instantiates B at t. This theory holds that the correlation between mental properties and brain properties are not accidental but law-like, and they are discovered in an empirical

characterized by the thesis of mind-brain supervenience, which is usually expressed by the claim that there can be no difference at the level of the mental without any difference in the brain (see the indiscernibility formulation of supervenience in section 3). In other words, if we find that two individuals have distinct mental states, then if physicalism is true, we will expect that there must be corresponding changes in the individuals' brain states. This, however, seems to clash with externalism. For *ex hypothesi*, the protagonists of the twin earth story are exactly alike with respect to their intrinsic physical properties and hence also their brain properties.²⁵ But if externalism is right, then it sounds as if there can be a difference in intentional properties without a corresponding change of brain state. This view seems to allow that intentional properties can '*float free*' from the physical and thus might seem like a kind of return to dualism.

A little reflection, however, will reveal that this objection is based on an implausible conception of physicalism. To avoid dualism, we need not stipulate so strictly that the supervenient bases of intentional properties comprise only intrinsic properties, or physical properties *within* the boundary of individuals' bodies. All that

way. It is generally agreed that construed as a theory about what mental states of the same type have in common, the type identity theory fails, since it is incompatible with one generally accepted view that mental properties (and indeed other non-basic (properties other than those posited by fundamental physics) are multiply realizable. It roughly asserts that for any mental property M, there could be more than one brain property instantiating M (see, e.g., Putnam (1967)). The classical statement of the identity theory can be found in Smart (1959).

²⁵ It is generally agreed that brain properties supervene on intrinsic physical properties.

that is required is that the supervenient bases of intentional properties all be physical facts. So, providing that relational properties are also physical facts (as indeed they are), externalism is perfectly consistent with physicalism.²⁶

7. The Common Concept Strategy Objection to Externalism.

Now let's return to the externalism argument. One main assumption on which the argument for externalism rests is that a given term used by the physical twins in the Twin-Earth story refers to different things (premise 1). Some philosophers, however, find this assumption doubtful, since it is not the case that our semantic intuitions *must* take the terms in question to refer to different things. To illustrate, consider the natural kind term or substance word "water". The objectors ask, what is wrong with saying that by "water" both speakers mean the same thing? For after all, the liquid filling the lakes and rivers on twin earth, though having a chemical structure different from H₂O, is functionally and observably completely indistinguishable from that on earth. Given this, why can't we take the extensions or references of "water" uttered by Oscar and his twin to be the same – both are substances having the properties of being colourless, tasteless, able to quench thirst, and so on. As Tim Crane puts it:

At the very least, it can be argued that our intuitive linguistic judgements do not settle the question whether XYZ is water; if, for example, it were discovered that what we pre-theoretically regard as water had many different underlying microstructures, maybe

²⁶ See, Botterill and Carruthers (1999): p.148.

three or four, maybe more, what should we conclude? It is implausible to suppose that our linguistic practices must dictate an answer *a priori*, so that we already know now what we would say about this case (Crane's emphasis) ((2002): p.123).

So it seems that one can say instead that both H₂O and XYZ are water, but have different microstructures.²⁷ Indeed, we actually call some substances superficially similar to H₂O but with different chemical structure "water," such as the chemical compound heavy water, which has the molecular formula D₂O.

Perhaps some may accept this objection, but point out that it has force only in some cases, where the word in question is *not* a *real* natural kind term, of which "water" is an example. But when we come across a real natural kind term "N" and discover that the substances it denotes are discernible with regard to their underlying microstructures – that is, what we pre-theoretically regard as N indeed have many different microstructures – we should dub them with different names, and say that only some are really N.

But what kinds of terms are really natural kind terms? Of course we cannot say that whenever we find a situation in which we individuate the microstructure of a substance rather than its observable features, then the term denoting the substance is a real natural kind term, for surely this definition is circular. Some may propose names of chemical elements as one kind of the candidate. However, this proposal fails, since anyone who studied chemistry will remember that some elements have isotopes –

²⁷ Objections of this kind can be found in Mellor (1977) and T. Crane (1991).

elements having the same number of protons yet different numbers of neutrons. Differing in number of neutrons is certainly differing in microstructure. But we still call the isotopes of an element by the name for that element – for instance, all isotopes of chlorine are called “chlorine.” It seems very hard to find an informative and non-circular definition of “natural kind terms.”

Moreover, it should be plausible to suppose that one of the central goals of science is to discover the microstructures of things. This, accompanied by the claim that the internal structures of things determine what they are, seems to suggest that science must have the last word in determining what words mean. But isn't this a form of scientific chauvinism? This point is nicely put by McCulloch:

It is natural to suppose that the principle underlying [Putnam's] strategy is that where there is a difference discerned by the relevant science – chemistry in this case – there we have to agree that there is an ultimate, or absolute difference. And this might seem unwarrantably deferential to science ((1995): p.172).

Suppose that feeling regular sharp headache that I long ago learned to call “pounding headache,” I decide to see a doctor. The doctor tells me that I don't have a pounding headache at all, for my brain does not show the pattern that pounding headache sufferers must exhibit – that is, the neurophysiological configuration of pounding headache. But is she justified to say this? Some English speakers may say No, since they may use “pounding headache” to refer to a particular kind of feeling or

qualia, regardless of its microstructural constitution.²⁸

We can summarize the objections just discussed as follows. It is not the case that our practices with substance words *necessarily* conform to Putnam's description – that is, it is not the case that when we find that instances of N, though superficially exactly alike, are discernible with regard to their microstructures, we inevitably judge them to be different things and thereby assign them different names. Indeed, there actually are some cases in which we do not individuate substances by their underlying structures (e.g., isotopes).

Plausible as this objection may sound, this common concept strategy is misguided. Establishing semantic externalism with regard to natural kind terms need not show that our use of substance words *must* follow the pattern Putnam describes. What is required is just that there *could* be a case in which our practice of using substance words meshes with Putnam's examples. That is, we can imagine a *possible* situation in which (1) we have a term "N" denoting a kind of thing N, where the distinction can be made between its appearance and underlying microstructure, and (2) we discover that not all instances of N have the same underlying microstructure, and (3) we judge that only some are N and thereby call them "N." If conditions (1) – (3) are all satisfied, then the truth of this kind of semantic externalism is established. Assuming the truth

²⁸ I borrow this example from G. McCulloch ((1995): p.173)

of premises 2 and 3, externalism with regard to intentional states is also established.

Psychology Argument

As I pointed out in the Introduction and the previous chapter, the externalist's argument works only if individualism is a plausible conception of scientific psychology. In this chapter I shall examine an argument for individualism that I shall call "the aim of psychology argument". I call it this because it derives individualism mainly from a conception of the aim of psychology, namely that psychology aims to concern itself only with explaining and predicting an organism's *raw* behaviour.

The structure of this chapter is as follows. In section 1 I shall provide a sketch of the argument (as below) (inspired by Stich (1978, 1983)¹⁷) and point out that the argument as it stands does not work, for one of its premises is false. In section 2 and 3 I shall explain how, by drawing a distinction between *autonomous* and *non-autonomous*, Stich modifies his argument and why he thinks that the truth of individualism provides a route to the externalist's conclusion. In section 4 I shall point out that Stich has no *a priori* argument for the claim that *autonomous* concern itself only with autonomous phenomena. This is the claim supported by the other premises of psychology. In sections 5 and 6 I shall point out that one is not so much against Stich's Psychology Argument as one is against the *a priori* argument for individualism. It may be regarded as an *a priori* argument for individualism.

¹⁷ See also Stich (1984) regarding the argument for individualism.

Chapter 3: Stich's Argument for Individualism: The Aim of Psychology Argument

As I pointed out in the Introduction and the previous chapter, the eliminativist's argument works only if individualism is a plausible constraint on scientific psychology. In this chapter I shall examine an argument for individualism that I shall call "the aim of psychology argument". I call it this because it derives individualism mainly from a conception of the aim of psychology, namely that psychology ought to concern itself only with explaining and predicting autonomous or narrow behaviour.

The structure of this chapter is as follows. In section 1 I shall provide a sketch of the argument from below presented by Stich (1978, 1983)²⁹ and point out that the argument as it stands does not work, for one of its premises is false. In sections 2 and 3 I shall explain how by drawing a distinction between autonomous and wide behaviour, Stich modifies his argument and why he thinks that the truth of individualism provides a route to the eliminativist's conclusion. In section 4 I shall point out that Stich has no *a priori* argument for the claim that psychology must concern itself only with autonomous behaviour. Nor is the claim supported by the actual practices of psychology. In sections 5 and 6, I shall point out that one argument implicit in Fodor's *Psychosemantics* may be regarded as an *a priori* argument for this

²⁹ Note that Stich (1996) repudiates this argument for individualism.

claim. However, the argument fails, since Fodor confuses causation with individuation. Finally, I shall argue that Stich's argument is invalid.

1. Stich's Argument for Individualism

Stich espouses individualism on the basis of the principle of autonomy, according to which "the states and processes that ought to be of concern to the psychologist are those that supervene on the current, internal, physical states of the organism." To put it in another way, "Any differences between organisms which do not manifest themselves as differences in their current, internal, physical states ought to be ignored by a psychological theory" (1983: p.164; see also (1978))

Stich calls properties that supervene on the current, internal, physical states of the organism "autonomous properties." The autonomy principle is intended to provide us with a normative constraint on how explanatory categories in scientific psychology should be individuated; that is, how mental states should be typed for the purposes of constructing scientific psychological theories.

But why should psychology respect this principle? To establish the truth of his version of individualism, Stich devises a now often-discussed thought experiment that he calls the "replacement argument". Suppose that a scientist succeeded in building a duplication machine, which can record a person's momentary internal physical states.

Now suppose also that you are placed in a duplication machine. After the scientist records all of your intrinsic physical features at a particular time t , a physical replica (or “replacement”, as Stich calls it) of you is produced. Stich claims that since you and your replacement share the same momentary autonomous states, the two of you will behave in the same way in all actual and possible situations.

Now if we agree that “psychology is the science which aspires to explain behaviour,” argues Stich, then it should be plausible to conclude that in articulating theories to explain human behaviour, psychology should be concerned only with autonomous properties. Or, equivalently, “any states or processes or properties which are not shared by [you] and [your] identically behaving replica must surely be irrelevant to psychology” (1983, pp.155-56). In short, psychology should respect the autonomy principle. We can reconstruct the above argument for individualism as follows:

1. Any two individuals who are indiscernible with regard to autonomous properties (e.g., you and your replacement) *must* behave in the same way in all circumstances.
2. Scientific psychology aims to explain actual and possible behaviour – that is, how an object will behave in various actual and hypothetical situations.

3. Thus any states, properties, or processes not shared by autonomously identical individuals should be omitted by psychology.
4. Thus psychology should concern itself only with autonomous properties.

Stich admits that though “there is an important kernel of truth” (1983, p.166) in the replacement argument, as it stands, it fails to establish the autonomy principle because premise 1 is false. It seems easy to conceive of situations in which two individuals are autonomously identical, yet do not behave in the same way. To see this, consider one such case provided by Stich (1983, p.166). Suppose that you own a car and are going to sell it to your friend. Unfortunately, after being duplicated, you are kidnapped and instead of you, your replacement goes to sell the car. Unaware of the fact that you have been replaced by a replica, your friend signs the documents and the transaction is successfully completed. But in this situation, your replacement’s behaviour, such as, signing the documents, receiving the cheque, and so forth, cannot be rightly described as selling the car, since she did not own it in the first place. Similarly, the following provides a counterexample to premise 1. Suppose that before being duplicated you promised Ann to help her complete a term paper and that after the duplication you died and your replacement goes to help Ann instead. Now even if your replacement helps Ann exactly as you would, were you there, your replacement’s behaviour could

not properly be described as fulfilling your promise, since it did not make a promise to Ann before.

2 Narrow and Wide Behaviour

To refine the argument, Stich distinguishes between two kinds of behavioural description: *autonomous* behavioural and *wide* behavioural description.³⁰ The former construes behaviour just as definite spatially-temporally defined movement of the body. Behaviour so described thus can be specified in purely physical or geometrical terms. Opening one's mouth, moving one's arms at so many meters per second, and contracting the biceps all can be regarded as examples of autonomous behaviour. Adopting wide behavioural descriptions means that behaviour is not just construed as physical movement but taxonomized in a way that takes account of the context in which bodily motions occur. Though how to characterize "context" precisely is a hard question, the following rough characterization seems to be on the right track. Instead of seeing a context as a particular like a space-time region, we can think of a context as a set of properties of an individual's environment and history. A physical behaviour is anchored in a context if and only if it bears a certain appropriate relation to the set

³⁰ In the current literature on philosophy of mind, wide causal power so characterized is often called "action" and "intentionally characterized behaviour".

of properties characterized by that context.³¹ A wide behavioural description construes behaviour as a hybrid relational property, in that the identity of a given wide behaviour is fixed by two factors: a certain bodily movement and a particular context. Opening the fridge, turning on the lamp, and making a cake all count as wide behaviours, since an individual can instantiate them only if her environment is amenable – for example, one can instantiate the behavioural type of opening the fridge only if she is in an environment that contains a fridge. Selling a car and fulfilling one's promise are also two examples of wide behavioural types, since tokens of them need to instantiate not only certain bodily movement types but also certain historical facts.

It is not hard to see that tokens of the same type of autonomous behaviour may count as distinct types of wide behaviour in different contexts. For instance, adopting autonomous behavioral descriptions, Oscar's and Oscar*'s verbal tokening of "bring me water" are behaviours of the same type (they consist of a particular form of vibration of the vocal cords). But they are distinct types of wide behaviour, since they are anchored in different contexts. Moreover, tokens of different types of autonomous behaviour may be counted as the *same* type of wide behaviour in the *same* context. Instances of the wide behavioural type selling a car may not all possess all the same

³¹ This account of context is similar to that provided by D. M. Walsh (1999 p. 633). A more detailed account of how context enters into the fixation of behaviour can be found in Enc (1995).

bodily movement types. For example, two sets of tokens of distinct types of bodily movements can be instances of selling a car, if the individuals that engage in the movements possess certain histories, namely owning a car before.

3. Refining the Argument

It is surely true, Stich admits, that under a wide or commonsense taxonomy of behaviour, which individuates behaviour in such a way that it takes account of the causal histories of individuals, as the above counterexample to premise 1 does, in some situations two autonomously identical individuals will be described as instantiating different behavioural types. Be that as it may, it seems plausible to expect that two autonomously identical individuals will engage in the same autonomous behaviour. Thus premise 1 of the above argument can be refined as follows:

- 1*. Any two individuals who are indiscernible with regard to autonomous properties (e.g., you and your replacement) *must* behave in the same way in all circumstances, where behaviour is understood as autonomous behaviour.

Stich then goes on to point out that there is no reason to require that psychology must adopt the wide or commonsense conception of behaviour. Like other scientific

disciplines, in explaining the phenomena that fall within its scope, psychology should devise its proprietary taxonomies. And so we should not expect that behaviour under all imagined descriptions must be explained by psychology. "Rather, the psychologist must select or formulate an appropriate descriptive language for his explananda. And the formulation of such a vocabulary will be a fundamental part of psychological theory construction" ((1983): p.167). The behavioural description that psychology *should* adopt, according to Stich, is the autonomous behavioural description. Premise 2 thus can be modified as follows:

2*. Scientific psychology ought to concern itself *only* with explaining (actual and possible) behaviour described autonomously.

According to Stich, psychology should be expected to explain and predict the occurrences of autonomous behaviour but not wide behaviour. Behaviour is generally regarded as a causal consequence of intentional states. So, in offering explanations and predictions of behaviour, psychologists need to identify the *nomic* mental causes (intentional states) of behaviour and formulate causal generalizations in terms of those causes. And though individuals can form their particular intentional states in a variety of ways, "it is their *being in those states* but *not how they come to have them* that is

relevant to their subsequent behaviour” (Wilson (1995): p.7) (my emphasis). As a result, in devising psychological generalizations, psychologists should ignore states not shared by individuals and their physical replicas, since it suffices to explain autonomous behaviour solely by properties ‘in the head’ – those properties that supervene on intrinsic physical ones.

The following case (see diagram (a))³² may help explain why Stich thinks that non-autonomous properties are causally irrelevant to autonomous behaviour.

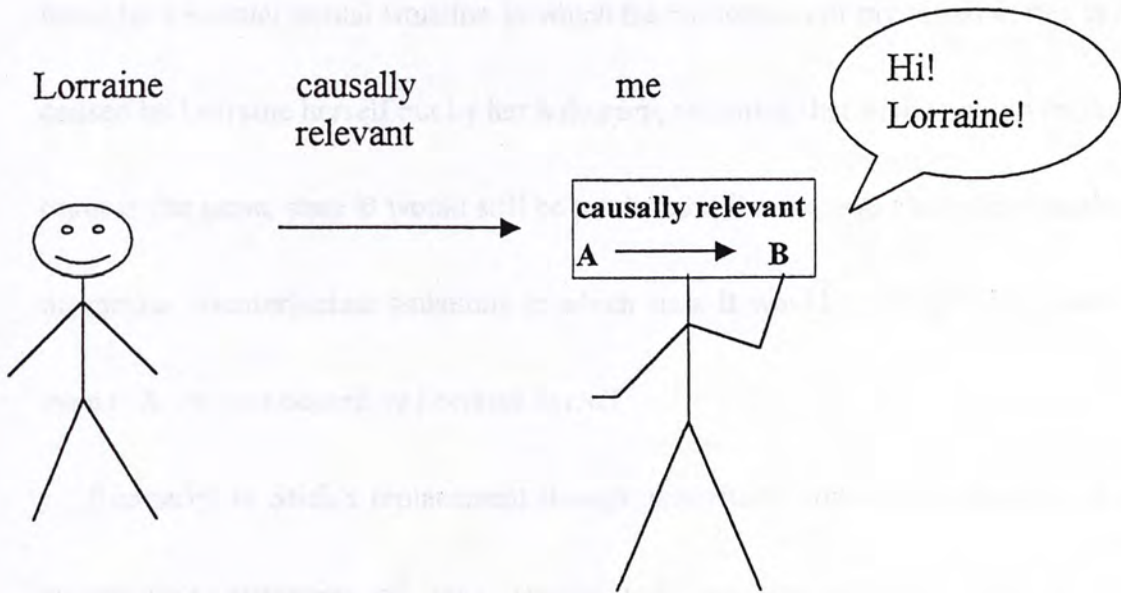


Diagram (b)

Suppose I see Lorraine. She is causally relevant to my producing state **A**, a stimulation of my visual cortex. After **A** interacts with certain information stored in my memory, I recognize her and state **B**, my thought that the girl standing there is Lorraine, is produced. State **B**, in turn, brings about a certain behaviour, say,

³² The diagram is borrowed with minor modifications from K. Sterelny (1990, p. 34), though he uses it to demonstrate a slightly different point.

producing the noise "*Hi, Lorraine*" by vibrating my vocal cord, waving my hand, smiling, and so on. In this case, we can say that state **A** is causally relevant to state **B**, that is, my having state **A** is partly caused by my mental representation of Lorraine. But it seems that the external condition (the presence of Lorraine) that causes me to be in state **A**, the causal ancestry of state **A**, is causally potent neither to my being in state **B** nor to my autonomously behavioural outputs, in that **B** and the behaviours it causes would be produced irrespective of **A**'s actual casual ancestry. For example, consider a counterfactual situation in which the stimulation of my visual cortex is not caused by Lorraine herself but by her hologram, assuming that what is going on in my brain is the same, state **B** would still be produced. Likewise, we can indeed conceive of various counterfactual situations in which state **B** would be triggered by state **A**, even if **A** was not caused by Lorraine herself.

Similarly, in Stich's replacement thought experiment, the causal ancestry of the autonomous properties of your replica and you are different: whereas your replacement's were produced by the scientist, yours were gradually formed by your interactions with your environment. But if explaining behaviour is the only concern of psychology, the causal ancestry of your and your replacement's autonomous properties "ought to be ignored by a psychological theory" ((1983): p.164).

As Stich points out, if individualism is true, it poses an eliminativist threat to

folk psychology. As we saw in the Introduction, folk psychology includes a great deal of nomological generalizations couched in terms of the contents of intentional states. And as I pointed out in the previous chapter, according to externalism, to a large extent, contents are non-autonomous properties, in the sense that they do not supervene on intrinsic physical properties. But if Stich's argument successfully establishes that the properties involved in the generalizations of scientific psychology must be autonomous, it follows that intentional states, because they are non-autonomous, cannot legitimately be employed by scientific psychology. As a result, folk psychology cannot be a part of scientific psychology. This, together with the claim that science is the measure of all things – only entities and properties posited by the sciences are real or genuine – leads us to eliminativism about folk psychology.

Stich not only agrees with the externalists that the contents of intentional states posited by folk psychology are non-individualist, but cheerfully accepts the conclusion of the eliminativist's argument. One of the main reasons he does so is that he holds that psychologists can do their work without making use of the notion of content.³³ However, not all philosophers think that eliminativism is tolerable. As a result, those who accept the conclusion of externalism, yet agree that individualism is

³³ Stich's view of this is supported by his syntactic theory of mind, a theory intermediate between neurophysiology and intentional psychology.

a proper constraint on scientific psychology, try to save folk psychology by developing a notion of content that is autonomous. In the next chapter I shall discuss one such attempt, Fodor's account of narrow content.

4. *Is The Replacement Argument Successful in Establishing Individualism?*

Stich thinks that his argument for individualism is a sound one. But is it? Let's first focus on premise 2*. Note that 2* is in fact a conjunction of two separate doctrines, as follows:

2(a)* Scientific psychology should explain *nothing but* behaviour, that is, the only explananda of psychology should be behaviour, and

2(b)* Behaviours explained by psychology must be described autonomously.

Now one might doubt that psychology should obey this normative constraint. We can agree with Stich that autonomous behaviour is among the explananda of psychology.

But why should psychology aspire to explain *nothing but* autonomous behaviour?

Isn't this constraint too restrictive? One way to evaluate whether or not a normative constraint on a given discipline is a plausible one, I think, is to examine whether that

discipline as it is actually practiced conforms to it. If it fails to do so, the plausibility of the constraint in question will be undermined. So, if we find that contrary to what Stich expects, psychology is *not* just concerned with explaining autonomous behaviour, then his argument for individualism in psychology cannot be sound because one of its premises is false.

Unfortunately, construed as an actual claim, 2 (a)* seems to be false simply because the explananda of some branches of psychology are not behaviour at all. As Eckardt points out, the central aim of cognitive psychology is to explain cognitive or mental capacities, such as language understanding and problem solving in logic, rather than behaviour. She writes that:

The explanandum of cognitive science is not exclusively autonomous behavior for the simple reason that most of the time it is not behaviour at all, autonomous or otherwise. Rather, *the basic aim of cognitive sciences is to explain the human cognitive capacities – what they are, how they are exercised, in virtue of what we have them, and how they interact* (my emphasis) ((1993): p.258).

Of course, as Eckardt goes on to point out, the explananda of cognitive psychology sometimes are behaviour, in the sense that exercises of some cognitive capacities involve behaviour (e.g., speech production). However, this fact cannot save 2(a)*, since it is surely untrue that the explananda of cognitive psychology are nothing but behaviour.

Another way to evaluate whether a normative constraint is a plausible one is to examine whether it is supported by *a priori* grounds – whether it can be derived from

certain plausible metaphysical doctrines. However, to my knowledge, Stich provides no such argument. Perhaps an argument implicit in Fodor's *Psychosemantics* may be used to support premise 2 (b)*.³⁴ Let's turn to it now.

5. Fodor's Argument for Premise 2 *: Narrow Behaviour and Crazy Causal Mechanisms

Fodor's argument, which has the structure of a *reductio ad absurdum*, can be roughly summarized thus. Psychology that adopts wide behavioural descriptions is committed to postulating impossible causal laws. Since such laws should be avoided, psychology should not adopt wide behavioural descriptions.

It is generally agreed by philosophers that behaviour is decisive in determining whether or not two individuals' intentional states agree in their causal powers.³⁵ If two individuals instantiate the same *type* of behaviour in all actual and possible circumstances, then the intentional states of the two individuals agree in their causal powers. To put it in a slogan, "[s]ame type of behaviour, same causal powers; different types of behaviour, different causal powers" (Macdonald (1995), p.164). And adopting wide behavioural descriptions implies that the intentional states of physical twins do not agree in their causal powers, since, as we have seen in section 3,

³⁴ The argument in Fodor (1991) may be seen as another one for premise 2*.

³⁵ Crudely, to say that an object has a particular causal power is to say that it can do or cause some effect in certain circumstances. The notion of causal powers will be explained in the next chapter.

if behaviours are described widely rather than autonomously, there will be some situations in which two autonomously identical individuals (physical twins) behave differently. By the behavioural criterion for the sameness of the causal powers of mental states, it follows that the physical twins' intentional states are distinct in their causal powers. And so we have a case in which two individuals share all the same autonomous properties, and yet they differ in causal powers. This violates the following doctrine, metaphysical individualism.³⁶

(MI): The causal powers of anything supervene on the intrinsic physical properties of that thing.

However, according to Fodor, violating **(MI)**, that is, denying that physical duplicates agreeing in all the same causal powers, requires us to postulate causal mechanisms that he thinks mad, crazy, or impossible.

To illustrate this point, Fodor invites us to consider the following case. Call a particle an "**H**-particle at time t " iff a particular dime of Fodor's is heads up and a particle a "**T**-particle at time t " iff the dime is tails up (1987: p.33). And call the possible world in which the orientation of Fodor's dime is heads up "**H** world" and

³⁶ A more detailed account of the nature of content of this principle will be explained in the next chapter.

that in which it is tails up “T-world” respectively. Then Fodor can change every particle in the universe from H-particle to T-particle and back again by merely turning over his dime. Do H-particles and T-particles agree in their causal powers? That is to ask, in taxonomizing explanantia of causal explanations of the behaviour of these particles, should we conform to (MI)? Fodor argues that the answer *must* be *Yes*. For *ex hypothesis*, both H and T particles are exactly alike with regard to their intrinsic physical properties. If we type-distinguish their explanantia in explaining their behaviour, that is, regard them as disagreeing in causal powers, this means that there could be differences in the causal powers of particles without differences in their intrinsic physical properties. This implies that the differences in causal powers are due to things *external* to the particles. And since the *only* difference between them is their contexts – H-particles are in H world, T-particles are in T world – the change in causal powers must be *caused* by their environments – specified by the orientation of Fodor’s dime. This means that there are causal mechanisms or laws that mediate the dependency between the orientation of Fodor’s dime and the particles’ causal powers *without any change in the particles’ intrinsic physical properties*. But such causal mechanisms, according to Fodor, are crazy.

Perhaps what grounds Fodor’s claim that violating (MI) will involve postulating crazy causal mechanisms is a conviction about causation – that causation must be

‘local,’ in the sense that for any physically constituted event to cause another one that occurs at a considerable distance, the cause must trigger a causal chain in such a way that each physically constituted event within the chain must be contiguous to its immediate cause and to its immediate effect.³⁷ This view is generally accepted by philosophers because denying it seems to commit us to a case of action at a distance.

Seeing physical twins as differing in the causal powers of their intentional states is analogous to the H-T particle case in respect to taxonomic practices. Since the physical twins, *ex hypothesis*, are autonomously identical, the difference in the causal powers of their intentional states must be explained by postulating *crazy causal mechanisms* – causal laws that enable a subject’s environment to *affect* the causal powers of her intentional states *without affecting her physiology*. However, Fodor thinks that:

there is no such mechanism; you can’t affect the causal powers of a person’s mental states without affecting his physiology. That is not a conceptual claim or a metaphysical claim, of course. It is a contingent fact about how God made the world. God made the world such that the mechanisms by which environmental variables affect organic behaviours run via their effects on the organism’s nervous system, or so, at least, all the physiologists I know assure me. (1987: p.40)

To avoid postulating such crazy causal mechanisms, Fodor thus concludes that psychology should employ narrow or autonomous behavioural descriptions rather than wide ones. For clarity, let’s reconstruct his argument as follows:

³⁷ See Cain (2002): p.164.

1. If psychology adopts wide behavioural descriptions, then there must be some situations in which the behaviour (tokens) of physical twins are regarded as *type distinct*.
2. If there are some situations in which the behaviours of physical twins do not belong to the same type, then their intentional states do not agree in their causal powers.
3. So, if wide behavioural descriptions are adopted, then the intentional states of physical twins do not agree in their causal powers.
4. That physical twins differ in the causal powers of their intentional states, however, implies that there are *causal* laws that mediate between the environments and the causal powers of twins.
5. So, the laws that correlate the twins' environments and the causal powers of their intentional states must be ones that can cause some changes in their causal powers without affecting their intrinsic physical properties (*since ex hypothesis*, physical twins are identical with respect to their intrinsic physical properties). Such causal laws are crazy or mad or impossible.
6. So, if postulating impossible causal laws is to be avoided, psychology should not adopt wide behavioural descriptions.
7. So, the only behavioural descriptions that psychology ought to adopt are

autonomous or narrow behavioural descriptions.

6. *Causal vs. Non-Causal (Constitutive) powers*

Premise 1 seems to be true. And assuming the behavioural criterion for causal powers of intentional states, premise 2 is true as well. However, is it true that admitting that the causal powers of intentional states of an individual can be changed by its environment without changes to the individual's physiological properties (autonomous properties) inevitably commits to causal mechanisms that Fodor thinks are crazy? Probably not, for there seem to be two different ways in which an object can acquire causal powers from its environment.

It is surely true that in some cases an object *acquires* causal powers from its environment (or that the environment *affects* or changes the causal powers of an object) because of causal laws that mediate the dependency between the object and its environment. For example, that an iron bar acquires the causal power of being magnetic in virtue of its proximity to a magnet is explained by *causal* laws that involve changes in the micro-composition of the bar. It is the bar's environment (its proximity to a magnet) that *causes* the iron bar to have the causal power of being magnetic.³⁸ Another example is provided by Fodor's passage quoted above. The

³⁸ See Cain (2002): p.162.

effect of an individual's environment on his neurophysiological properties is explained by causal laws that involve changes in the individual's intrinsic physical properties. Be that as it may, it is not the case that an object *must* acquire causal powers from its environment in this way. An object can acquire a particular causal power by being embedded in an appropriate context. In other words, causal powers of this kind are *constitutive* rather than causal. Call this kind of causal powers "wide causal powers".

So, in saying that the environment changes or makes a difference to the causal powers of the intentional states of the physical twins, one is not committed to the claim that such changes are *caused* by the individual's environment. As Egan put this point:³⁹

Fodor's argument, however, simply misconstrues the anti-individualist's position. The proper construal of the anti-individualist claim is not that differences in the environments *cause* a difference in the twin's mental states, but rather that environmental (contextual) factors are relevant to the *type-individuation* of their mental states. Our theories, and the taxonomies employed by them, are not caused by the world in any direct sense. In claiming that psychological taxonomy is sensitive to contextual factors the anti-individualist is not postulating any mysterious causal processes (Egan's emphasis) ((1991): p.189).

Perhaps Fodor can deny the legitimacy of wide causal powers by arguing that wide causal powers (causal powers that are not causal but constitutive) are *not* real or genuine causal powers⁴⁰ or that wide causal powers should not be recognized by

³⁹ Similar objections can be found in Burge (1986) and Cain (2002).

⁴⁰ Perhaps it may be little strange to say that wide causal powers are not causal. To ease this worry, we may delete the word "causal", just calling it "wide powers".

'real' sciences. However, Fodor does not argue for this. To be sure, in many scientific disciplines causal powers are not wide (e.g., magnetism, solubility, and c-fiber activation). "But nothing about science in general can be read off from this fact" (Cain (2002): p.163). Indeed, as we shall see in the next chapter, wide causal powers are prevalent in the taxonomy of explanatory kinds in various scientific disciplines.

If the above reasoning is right, then premise 4 is false. And so Fodor's argument for 2* is unsound.

7. Conclusion: Stich's Unsuccessful Argument for The Principle of Autonomy

Even if we grant that the only aim of psychology is to explain autonomous behaviour, does Stich successfully rule out all non-autonomous or relational properties as illegitimate to psychology? It seems not. Comparing Stich's replacement thought experiment with Putnam's and Burge's may help to illustrate this point. In all of these thought experiments the autonomous properties of the protagonists, *ex hypothesis*, are identical; what varies is just one kind of non-autonomous property. Whereas in Putnam's thought experiment, it is the natural environment in which the individuals live that varies, in Burge's, the only difference between the individuals is their linguistic and social environment. But in the scenario Stich describes, while the causal histories of the autonomous properties of the physically identical twins vary, other

non-autonomous properties related to their natural, linguistic, social, and cultural environment remain wholly unchanged. So the argument at most shows that one kind of non-autonomous property – historical properties or the causal ancestry of autonomous properties – is redundant and hence explanatorily irrelevant to psychological explanations. It remains silent on the question of whether other kinds of non-autonomous properties, such as natural and social environmental properties, are irrelevant.⁴¹ And so it fails to establish the claim that psychology should include only autonomous properties as proper scientific explanatory kinds.⁴²

Do individualists have other routes to their doctrine? In the next chapter I shall discuss another argument for individualism provided by Fodor, which derives individualism mainly from a doctrine concerning how the explanantia of psychology should be individuated.

⁴¹ As I shall show in the next chapter, there is indeed a sense in which natural environmental facts are relevant to explaining autonomous behaviour.

⁴² See Eckert (1993): p. 156.

Chapter 4: Fodor's Argument for Individualism: The Explanan Argument

In the previous chapter, I examined an argument for individualism from a normative claim about how the explananda of psychology should be taxonomized. I concluded that it fails because there is no cogent reason to require psychology to concern itself only with narrow or autonomous behaviour.

Other than arguing for individualism from the aim of psychology, one can argue for individualism from how the explanantia of psychology should be individuated (I call this type of argument "the explanan argument"⁴³). Fodor (1987: ch 2) presents one such argument, which can be summarized roughly as follows: explanatory categories or kinds devised by the sciences must be individuated by their causal powers. The causal powers are determined by or supervenient on intrinsic physical properties. So, if psychology is intended to be a scientific discipline, the explanatory kinds it employs must be autonomous properties, or those that supervene on intrinsic physical properties.

The structure of this chapter is as follows: In section 1 I shall provide a sketch of Fodor's argument from above. Then in section 2, I discuss Fodor's attempt to reconcile folk psychology and individualism by working out a different kind of

⁴³ Note that in the current literature on philosophy of mind this type of argument is usually called "*the argument from causal powers*" (e.g. Wilson (1995), Cain (2002)).

content: *narrow content*. In section 3 I shall argue that contrary to what Fodor thinks, global individualism cannot be derived simply from an uncontroversial view about how explanantia should be typed. Nor does it conform to the fact that explanantia in various sciences are relationally taxonomized. I will discuss various possible modifications of global individualism to account for this fact. However, I shall argue that none of them is successful, since such refinements turn Fodor's argument into an invalid one.

1. Fodor's Explanan Argument

According to Fodor, one of the central goals of the sciences is to explain phenomena by constructing causal explanations. In doing so, different scientific disciplines should devise their own proprietary behavioural descriptions (*explananda*) and *explanantia*. Fodor takes it for granted that in general causal explanations provided by the sciences involve a matter of subsuming events to be explained under certain projecting and confirming causal laws or generalizations.⁴⁴ Causal laws "subsume the things they apply to in virtue of the causal properties of the things they apply to" (Fodor (1987): p.34).⁴⁵ To explain why an event token *e1*, which belongs to event type *E1*, occurred, requires (typically) specifying a prior event token *e2*, which belongs to event type *E2*,

⁴⁴ See, e.g., (1987), p. 34, (1994), p. 3, (1998), p.7.

⁴⁵ The above passage indicates that Fodor adopts a Humean account of causation and Hempel's account of causal explanation.

and a causal generalization or law, the antecedent and consequent of which are **E1** and **E2**, respectively. For example, the event of a toy truck's accelerating 10 ms can be causally explained by subsuming it under Newton's second law and the fact it was acted on by a force of 20 newtons and has a mass of 10 kg.⁴⁶ Similarly, that an object has a certain kinetic energy is explained by its being subsumed under the law that $\text{kinetic energy} = 1/2 \times \text{mass} \times \text{velocity}^2$.

Fodor argues that in constructing causal generalizations, the sciences must taxonomize their explanatory kinds (their explanantia) in a way that respects the following methodological principle, which he calls "*global methodological individualism*":

(GI): Explanatory kinds or categories in the sciences must be individuated by their causal powers⁴⁷.

(GI) provides scientific disciplines with a normative guide to how things, states of things, and whatever, are classified: anything with the same causal powers should be regarded as belonging to the same kind. In other words, according to **(GI)** any

⁴⁶ More precisely, we should say that the event is explained by *an instance* of Newton's 2nd law of motion.

⁴⁷ "*Categorization in science is characteristically taxonomy by causal powers,*" as Fodor ((1987): p.44) puts it.

classification scheme that groups things under different kinds if they agree in causal powers must be counted as scientifically illegitimate. *Same causal powers, same scientific kinds*, to put (GI) in a slogan.

But *what is it for something to have a causal power, according to Fodor?* To a first approximation, to say that a thing (an individual, an object, an event, and the like) possesses a particular causal power is to say that it has a capacity to do something or cause certain effects in certain nomologically possible contexts.⁴⁸ Typically, causal powers are characterized by conditional or hypothetical statements having the form “If ... then.” For example, to say that a sugar cube has the causal power of being soluble in water (or the disposition to dissolve in water) is to say that if immersed in water, then it will dissolve. To claim that a mental state (token) has certain causal powers is to claim that it has the capacities to bring about certain effects, namely behaviours and other mental states, in certain contexts.

There are two main analyses of causal powers: *instrumental* and *realistic*. According to the former, to ascribe certain causal powers to a thing is *just* to cite the fact that certain true conditional statements hold for it. Thus, causal powers such as being soluble in water, on this view, are *hypothetical* or *conditional* properties of objects – dissolving if immersed in water. The latter, by contrast, takes causal powers

⁴⁸ Crudely, nomologically possible contexts or circumstances are those in which no natural laws are violated.

as *categorical internal states* that causally explain why certain conditional statements hold for the object.

One should note that causal power is a *counterfactual-involving* notion, in the sense that something can have a particular causal power even if it never exercises it. For example, a sugar cube can count as possessing the causal power of being soluble without ever exercising that power. All that is required is a disposition such that had it been immersed in water, it would have dissolved. As a result, in determining whether or not an object possesses a particular causal power, we need to consider how it behaves not only in actual cases but also in hypothetical situations. "You have to judge identity and difference of causal powers in a way that bears the counterfactuals in mind" ((1991): p.8), as Fodor reminds us.

With the notion of what it is for an object to have a causal power, Fodor can now formulate his proposed criterion of the identity of causal powers: "identity of causal powers is identity of causal consequences across nomologically possible contexts" (1987, p.44). More precisely, it can be formulated as follows:

Two objects *a*, *b* have the same causal powers just in case for all nomologically possible contexts or circumstances *C*, if *a* can cause certain effects in *C*, then so can object *b*, even if *b* in fact is not in *C*. Or, to put the same point alternatively, if two

objects possess all the same causal powers, then all true hypotheticals of the form “*a* can cause *E* in circumstance *C*,” if “*b*” is substituted for “*a*” will remain true and vice versa (see (1989)).

But what determines the causal powers that a thing has? Or, in virtue of what does something possess the causal powers it has? Fodor’s answer is that the causal powers of anything are *determined* or *fixed* by intrinsic physical properties (crudely, properties posited in theories of fundamental physics), or by those properties that supervene on intrinsic physical properties.⁴⁹ Since the notion of determination can be captured by that of supervenience (see chapter 2, section 3), Fodor’s claim thus can be put as:

(MI): The causal powers of anything supervene on the intrinsic physical properties of that thing: two objects sharing all their intrinsic physical properties must also share all of their causal powers across all nomological contexts.

It is easy to see that **(GI)**, together with **(MI)** jointly imply that the explanatory kinds adopted in the sciences, whatever they are, must be individualistic, that is, they must

⁴⁹ As Fodor puts it “*Causal powers supervene on local microstructure.*” ((1987): p.44).

be autonomous properties.

Now since Fodor thinks that explanations provided by scientific psychology must involve the construction of projectible and confirming causal generalizations⁵⁰ couched in terms of “mental states that are specified under intentional description: viz. among *mental states that are picked out by reference to their content*” (Fodor’s emphasis) ((1998): p.7), it follows that if two individuals’ intentional states agree in their causal powers, psychologists, for fear of violating (GI), must regard their states as belonging to the same type. Any taxonomy adopted in psychology that type-distinguishes intentional states that agree in their causal powers should be ruled out by (GI). Given (MI), it follows that psychology should concern itself only with autonomous properties. The above reasoning may be rather complicated, so for clarity we can reconstruct it as the following argument:

Fodor’s argument for individualism from above

1. Explanatory kinds devised by the sciences must be individuated by causal powers.

That is, if two things agree in their causal powers, any science must regard them as of the same kind.

2. The causal powers of anything are determined by or supervenient on the intrinsic

⁵⁰ Note that Fodor’s view is controversial. R. Cummins (1983), for example, argues that psychological explanations should be thought of as functional analysis rather than as nomological subsumption.

physical properties of that thing.

3. Psychology is intended to be a scientific discipline that provides causal explanations by constructing causal generalizations or laws.
4. So, explanatory kinds devised by psychology, namely intentional states, ought to be individuated by causal powers: if two intentional states agree in their causal powers, they must be counted as of the same psychological type.
5. So, explanatory kinds devised by psychology must be autonomous properties.

Except for premise 3, all of these premises are controversial. As a result, Fodor's argument is not generally accepted by philosophers. Some, for example, question the truth of premise 1. On the one hand, some point out that when investigating actual individuating practices in scientific disciplines such as evolutionary biology, economics, and anthropology, we find that they do not always taxonomize relevant entities and properties by their causal powers. And so, assuming that a plausible constraint on science must do justice to the actual practices of scientific disciplines, we should reject global individualism. On the other hand, some argue that contrary to what Fodor believes, (GI) is not supported on *a priori* grounds and so its plausibility is in doubt. Moreover, some argue that on some interpretations of "causal powers," premise 2 turns out to be false, since intrinsic physical properties are insufficient to

determine causal powers according to those interpretations. Worse, Wilson (1995) argues convincingly that there is no interpretation on which Fodor's argument is a sound one: either it rests on false premises, or it is invalid in that it equivocates on the notion of causal powers in premise 1 and 2. In the following sections I shall discuss these objections in turn.

At first glance, if one is an individualist about scientific psychology and agrees that folk psychology taxonomizes its explanatory kinds widely, then for fear of inconsistency, it seems that one must reject the posits of folk psychology as scientifically illegitimate. Assuming that only the posits of the sciences are real, eliminativism then follows immediately. As we have seen, Stich cheerfully welcomes the eliminativist's conclusion. Though like Stich, Fodor accepts both the view that folk psychology is non-individualistic and that scientific psychology must be individualistic, he is unwilling to accept eliminativism. Indeed, he thinks that:

If commonsense intentional psychology really were to collapse, that would be, beyond comparison, the greatest intellectual catastrophe in the history of our species; if we're wrong about the mind, then that's the wrongest we've ever been about anything. The collapse of the supernatural, for example, didn't compare. (1987: xii)

One main reason why Fodor thinks eliminativism is unbearable, I think, is that if it were true, we could not account for the successes of folk psychology in predicting and explaining human behaviour without appeal to miracles. His reasoning can be

regarded as an instance of the so-called No Miracle argument,⁵¹ a kind of argument in philosophy of science often appealed to by realists. In its crudest form, the argument goes like this. Given that a scientific theory (e.g., Maxwell's theory of electromagnetic fields) has been tremendously successful in prediction and explanation, if miracles are to be avoided, the best way to account for its success, is to endorse the ontology to which the theory is committed – that is, to take the theoretical posits of the theory to refer to genuine properties and to regard its claims as approximately true. Folk psychology, according to Fodor, is an implicit empirical theory with tremendous predictive and explanatory potency. So, if eliminativism were true, how could the successes of folk psychology be explained without appeal to miracles?

Second, Fodor does not believe that we can do psychology without the structure of intentional explanations, as he thinks that there are psychological generalizations that can only be captured or couched in terms of content. As he says, “the reliable explanatory generalizations of any psychology that we can now foresee will be intentional through and through; if there are no intentional laws, then there are no psychological explanations” ((1994): p.3).

However, Fodor thinks that any adequate taxonomy of scientific psychology must

⁵¹ See, e.g., (1987): ch 1.

respect individualism and agrees that the theoretical posits of folk psychology are not individualistic. So, in order to ‘have it both ways’ – to respect individualism and to save folk psychology – Fodor tries to revise wide intentional states in a manner that “renders them individualistic while preserving the basic structure of intentional explanation” (Stalnaker (1989); (1999): p. 170). How can he do this? Before getting into Fodor’s own account, it would be appropriate to clarify the general character of *revisionist* project⁵².

2. A Response to Individualism: Rendering Intentional States Individualistic

Even if our having a given (wide) intentional state (type) is in part dependent on something external to us, as externalists argue, no one denies that every intentional state has an internal component, which is wholly within the agent and partly determines wide intentional states. For after all, intentional states such as believing that water is wet, unlike the properties of being three miles away from Fung King Hey building and that of living in a watery world, are not *pure* relational properties – that is, not properties the possession of which is merely a matter of facts about the object’s environment. We can dump any object on Twin Earth and thereby give it the property of being in a world where there is twater in lakes, but we “cannot just dump an object

⁵² The following account of the revisionist’s project is from Stalnaker ((1989)).

on twin-earth, and thereby bring it to believe that twin-water quenches thirst” (Joe Lau (1994)).

So intentional states are *hybrid* relational properties, in the sense that each of them is composed of two elements: an internal component, a state that is internal to the subject, and an external component that consists of the relation holding between the subject and its environment. Revisionists aim to find some way to factor out the internal contribution to an intentional state – the internal component of the state that is supervenient on the subject’s intrinsic physical properties. We can call this component “*narrow content*.”

However, at least at first sight, the notion of narrow intentional states seems to be incoherent. For an intentional state is generally conceived to be a *relation* between the subject and a kind of object, namely a content. For example, the statement that Lorraine believes that her boyfriend is smart expresses the relation between an individual – Lorraine – and a particular content – *her boyfriend is smart*. So how can an intentional state be both a relational property (in the above sense) and a property that supervenes on intrinsic physical properties? “Isn’t it obvious that semantic properties, and intentional properties generally, are relational properties, properties defined in terms of relations between a speaker or agent and what he or she talks or thinks about? And isn’t it obvious that relations depend, in all but degenerate cases,

on more than the intrinsic properties of one of the things related?” (Stalnaker (1989); (1999): p.170).

The apparent incoherence disappears once we distinguish two senses in which a concept or property is *relational*. Consider mass and length, two properties usually cited as typical examples of intrinsic properties. We regard them as intrinsic because we expect that any two objects physically identical must have the same mass and length. Let's call this sense of a property being intrinsic "*ontologically intrinsic*." Properties that are not ontologically intrinsic are *ontological relational*. Weight is a typical example of an ontologically relational concept, in that two objects indiscernible with respect to autonomous properties may not share the same weight if placed in environments with different local gravity.⁵³

That a property is ontologically intrinsic, however, is perfectly compatible with saying that it is relational in another sense. To illustrate this point, let's consider again the physical magnitudes mass and length. Their intrinsic nature does not prevent us from *measuring, specifying, and picking them out* by things external to them, that is, by relational means. To say that a book has a mass of 3 Kg is to say that it *bears a particular relation* to the International Prototype Kilogram: it will balance on an equal-arm balance these objects each of which balances the Standard Kilogram. To

⁵³ Note that this sense of being intrinsic and relational is that I explain in chapter 2.

say that a stick has a length of 1 meter is to say that it has the same length with the Standard Meter. A relational predicate, together with a number, is usually used to specify and pick out the fact that an object has a property, for example “having a mass of 3 Kg.” Let’s call a property’s being relational in this sense “*semantically relational*.”⁵⁴

Now it should be clear that no incoherence lies in the notion of narrow intentional states. What revisionists want to work out is a kind of intentional state that is ontologically intrinsic but semantically relational: these states are autonomous properties but, like wide or ordinary intentional states, can be picked out or specified relationally.

It is of the utmost importance to know that the task before the revisionists is not just to factor out the internal contributions to a particular wide intentional state such as believing that water is wet. What is required rather is to demonstrate that corresponding to every wide intentional property, there can be *at least one* narrow counterpart and how the narrow intentional states *determine* the wide ones in various contexts. The idea can be captured by the following abstract schema (Stalnaker (1999): p. 24)

⁵⁴ Accounts of the distinction between these two senses of being relational can be founded in Stalnaker (1989), Kim (1996): ch. 8).

For any agent x , x believes that p iff $(\exists Q)(x \text{ believes}_n Q, \text{ and } C(Q, \text{ that } p))$, where “*believes_n*” expresses the relation holding between agents and narrow contents, the quantifier ranges over narrow contents, and the C relation relates narrow contents to their wide counterparts.

What revisionists need to do is to flesh out the schema by specifying (1) what it is for a subject to possess a narrow belief with a particular content p , and (2) what a context consists in, and (3) the *correlation* between wide intentional states and their narrow counterparts: how a narrow intentional state, accompanied by context (C), determines a wide one. The revisionist’s project counts as successful only when (1) to (3) are satisfied. Equipped with this overview of the revisionists’ project in general, let’s turn to Fodor’s own version of revisionism.

2.1 Fodor’s Account of Narrow content⁵⁵

Fodor’s own distinctive account of narrow content is closely related to his diagnosis of the twin-earth thought experiment. What makes Oscar and Oscar* have different intentional states, according to him, is the following fact: assuming semantic externalism for certain terms, the truth-conditions of utterances and tokens of

⁵⁵ Note that Fodor (1994) rejects his account of narrow content.

intentional properties such as beliefs that involve such terms are in part constituted by the environment in which the utterers and the believers are situated. As a result, the truth conditions or contents of their utterances and beliefs should be regarded as distinct because Oscar and his twin are embedded in different environments. To factor out the internal contributions or to subtract the environmental facts relevant to the determination of contents of utterances and thoughts – that is, to put truth conditions back into the head – Fodor proposes that we can construe the internal components of contents as functions (in the mathematical sense) that take possible contexts as arguments and truth conditions as values. In other words, narrow contents are functions that map possible contexts onto truth conditions. Since such content has no dealings with the environment and hence is wholly inside the believer, Fodor calls it “*narrow content.*”

So, to have an intentional state with a particular narrow content is to have a state with a particular function that enables an individual in various contexts to entertain various intentional states with wide content. The criterion of narrow content can thus be formulated as this: *the narrow contents of two tokens of intentional states (and of utterances) are identical just in case in all possible contexts they will generate the same truth conditions.*⁵⁶

⁵⁶ Fodor’s characterization of narrow content is inspired by D. Kaplan’s semantic account of context-dependent expressions such as demonstratives and indexicals (e.g. “This”, “That”, “I”, “You”,

Perhaps the idea can be made clearer by an example. The narrow content shared by Oscar and his twin when each of them says and believes that water is a boring drink is the function that when anchored on earth generates the truth condition that *water (H₂O) is wet*, and when embedded in twin-earth yields the truth condition that *twater (XYZ) is wet*.

2.2 Criticisms of Fodor's account of narrow content

As I have said, any adequate account of narrow content must satisfy conditions (1) to (3) in the abstract schema. So now we need to see whether Fodor provides us with an informative guide to how to individuate contexts. The following is his account of what a context is:

There is presumably something about the relation between Twin-Earth and Twin-me in virtue of which his 'water'-thought are about XYZ even though my water-thoughts are not. Call this condition that's satisfied by {Twin-Me, Twin-Earth} condition C (because it determines the Context of his 'water'-thoughts). Similarly, there must be something about the relation between me and Earth in virtue of which my water-thoughts are about H₂O even though my Twin's 'water' - thoughts are not. Call this condition that is satisfied by {me, Earth} condition C' ((1987): p.48).

We may agree with Fodor that there is something about the relation between the believer and her environment such that in virtue of standing in such a relation that her belief with a particular narrow content generates a belief with the corresponding wide

"Here", "There"). According to Kaplan, context-dependent terms have characters or linguistic meanings (e.g. the character of "I" is the utterer of the utterance), which yield truth conditions of utterances in particular context but are constant over different contexts of utterances.

content or truth condition. One can hardly deny this if he thinks that intentional states are hybrid relational properties. But what does this relation consist in?

Fodor proposes that a particular context should include anything relevant to the determination of wide content. But such a specification is uninformative, since context, *ex hypothesis*, is something external to the individual that is relevant to determining beliefs with wide content. And surely it is true that context so construed, together with narrow content construed as the internal contribution to wide content, can determine a particular wide content. But this assertion just amounts to restating the point that intentional states are hybrid relational properties.⁵⁷ It still remains unclear what is involved when an individual with a narrow content is embedded in a particular context.

The triviality of Fodor's characterization of context can be made more prominent by comparing it with the concept of weight. Physicists tell us that weight is a hybrid relational concept, since for any object to have a particular weight, other than instantiating a certain intrinsic property – having a particular mass – it must be embedded in a certain context, a particular local gravitational field – which serves as the external component that contributes to the object's having that weight. To have a particular mass is to instantiate a particular function that when embedded in different

⁵⁷ This objection is inspired by Stalnaker's (1989); (1999): pp.176-77).

local gravitational fields will determine a corresponding weight. In the case of weight, physicists have provided us with a concrete procedure about how the relevant contexts – gravitational fields – are to be individuated. We are not just told that the external component of having a particular weight is just anything external to the object and relevant to the determination of having that weight.

Moreover, Fodor claims that the narrow content of Oscar's and Oscar*'s belief that water is wet is the same. As a result, were Oscar transported to twin earth, the truth condition of his belief would immediately shift to be *twater (XYZ) is wet* (and the truth conditions of all of his intentional states expressed by using “water” would also shift analogously). Likewise, were he transported to the earth, the wide content of Oscar*'s belief expressed by the same type of utterance as Oscar would at once shift to be *water (H₂O) is wet*. In short, Fodor's account entails that whenever the contexts of the narrow contents of belief tokens and utterances involving natural kinds are changed, the wide contents of the beliefs will thereby be changed.⁵⁸ This, however, seems at odds with the standard interpretation of twin-earth story. For it is generally agreed that were Oscar transported to twin-earth, his beliefs expressed in terms of “water” would still be thoughts about water (H₂O). Similarly, Oscar*'s contents of utterances and beliefs would still remain unchanged even if he were transported to the

⁵⁸ This objection (though not the same as) is closely related to the one that denies changes in an individual's environment will cause changes in the causal powers of her intentional states. For more detailed discussions of the latter view, see Davies (1986) and Cain ((2002): p.163)

earth: when he believes that water is a boring drink on earth, what he believes is the same as what he believes on Twin Earth – that *XYZ is a boring drink*. In other words, the wide contents of such thoughts will *not* be affected by his *environment immediately*.⁵⁹

There are other serious objections to Fodor's account of narrow content. Stalnaker in fact argues that any attempt to complete the above schema will face a great deal of obstacles ((1990), (1993)).

Fodor needs to provide a theory of narrow content to save folk psychology because he holds that individualism in psychology is a plausible constraint. But is it? In the following sections I shall argue that Fodor's argument cannot be sound. I begin with his argument for global individualism.

3. Examining Global Individualism – Fodor's A Priori Argument

As I said in section 1, Fodor believes that global individualism can be merely derived from the fact that science aims at providing causal explanations for phenomena and that causal explanations involve subsuming the things to be explained under causal generalizations, or, equivalently, ascribing causal properties to the explananda. Since

⁵⁹ However, since the standard interpretation also claims that were Oscar lived in twin earth for a considerable period of time, since he has interacted with a lot of twater (XYZ), his thoughts expressed in terms of "water" would gradually become twater-thoughts. So, Fodor's account thus can do justice to the standard interpretation if we take account of temporal considerations.

the explanata of causal explanations or the antecedents of causal generalizations must be taxonomized in accordance with causal similarities and differences, Fodor concludes that they must be those *individuated by their causal powers*.

The main flaw in this argument is that the conclusion does not follow from the premises. For it is one thing to require that a classification scheme for causal properties assign distinct token phenomena (events, individuals, and whatever) to the same type just in case they agree in relevant causal similarities. But it is surely quite another thing to require that it must do so just in case they agree in causal powers. Grouping token phenomena under the same type in virtue of causal similarities may involve certain environmental and historical properties of those phenomena rather than what effects that those phenomena can bring about. That is to say, individuating causal properties by causal similarities permits them to be individuated perfectly well in a way that is *essentially backward looking* – what determines whether or not an object belongs to a category is *what it is caused by* rather than *what it can cause*. As

Wilson puts this point:

The relevant causal similarities between two phenomena in a given discipline may involve the causes of those phenomena or the causal relations they stand in, rather than what those phenomena are capable of causing. The historical and relational properties that two entities share may well explain why those entities share many other properties, and there is no reason to regard explanations citing such properties as non-causal (Wilson's emphasis) ((1995): p.33).

One such example is the property of being a victim of 9-11 incident. Whether a

person belongs to this category essentially hinges on her history or even her parent's history rather than what she can do or cause. This property can help explain many other properties of its instances. To explain why an individual's body was injured, it may be informative to cite the fact that she was a victim of the 9-11 incident. And it seems that explanations that make use of this property as explanans are not 'less causal'. In the next section I shall give some examples to show that in addition to examples from ordinary life, explanantia of causal explanations individuated by their causes are widespread in scientific disciplines, including natural and social sciences.

3.1 Counterexamples to Global Individualism

It seems reasonable to suppose that a normative constraint on science is plausible if it is countenanced on both *a priori* and *a posteriori* grounds – that is, if it can be derived from certain metaphysical assumptions and if it conforms to actual scientific practices. The previous section showed that the demand that causal properties be taxonomized by causal similarities does not imply that they must be taxonomized by causal powers. This shows that, contrary to what Fodor thinks, (GI), as a general constraint on taxonomizing explanantia of causal explanations, does not follow merely from *a priori* grounds – from the uncontroversial claim that explanantia of causal explanations must be individuated by causal similarities. This provides us with a

reason for doubting (GI).

Can (GI) be defended on *a posteriori* grounds? That is, do actual individuating practices in scientific disciplines conform to (GI)? Are causal properties *always* taxonomized by causal powers?

No doubt some scientific disciplines classify explanantia on the basis of their causal powers. Being soluble in water, a property figuring in certain causal generalizations in chemistry, for instance, is one such example, since what an object can do or cause constitutes the individuating criteria that determines whether or not it is soluble – to be so, the object must dissolve when immersed in water. Another example is the property of having a mass of 1 Kg. To specify the individuation-condition for it is to specify what an object can cause by virtue of it in various different contexts. For instance, in virtue of possessing a mass of 1 Kg, an object will accelerate with a certain value and will have a certain kinetic energy in certain circumstances.

As many philosophers have pointed out (e.g., Burge (1986), (1989); Egan (1991), Wilson (1995)), close examination of actual explanatory practices in a variety of sciences indicates that they do not solely or even primarily classify causal properties by their causal powers, relational taxonomy – that is, individuating antecedents of causal generalizations in a backward-looking way – is widespread in scientific disciplines such as biology, economics, linguistics, biology, and astronomy.

Consider the concept⁶⁰ of species, a central explanatory kind in evolutionary biology.⁶¹ Many biological generalizations are defined over it. Cope's law, for example, asserts that unspecialized species tend to avoid extinction longer than specialized species. According to the traditional understanding of species, a particular species such as *Canis familiaris* or *Drosophila melanogaster* is defined by its morphological or genetic properties. The former of these views holds that what determines whether an organism belongs to a particular species is essentially determined or fixed by a set of observable features – determinate forms of phenotypes. For a long time, “[t]he search for similarities and differences in phenotypic traits [was] at the heart of most classification of taxa in biology” (Caplan (1980)). The systematic procedures that decide whether a given domestic dog falls under the species *Canis familiaris*, for example, are based on whether it possesses certain phenotypic traits such as having four legs, a tail, upper and lower teeth, and so on. The latter of these views, by contrast, holds that the essence of a particular species is determined not by its phenotypic properties but by its genetic properties, for example, having particular sequences of DNA in the genome.⁶²

For a period of time, these two views dominated the taxonomy of species.

⁶⁰ Some philosophers distinguish concepts from properties, but here I take them to be the same.

⁶¹ That the concept of species is relationally taxonomized is pointed out by Egan (1991) and Wilson (1995).

⁶² In the current literature on philosophy of biology, these views on the taxonomy of species are called *essentialism* on species. For a detailed discussion of this view, see Ruse (1998) and Wilson (1999).

Nevertheless, nowadays most biologists agree that it is inappropriate to define species solely in terms of morphological or genetic properties. An appropriate definition of species, they propose, should involve not only the physical constitution of an organism but also its causal ancestry. Whether a group of organisms are regarded as belonging to the same species, according to these biologists, is determined not only by whether they possess certain morphological or genetic properties but also by whether they are descended from a common ancestor. Mayr, for instance, defines species as “ a reproductive community of populations (reproductively isolated from others) that occupies a specific niche in nature” ((1982): p.273). A consequence of this view is that two organisms may have the same physical constitution – namely the same morphological and genetic properties – but do not count as belonging to the same species. Even if biologists find a dog sharing all of the morphological and genetic properties of my dog Billy, if the two do not share a common evolutionary ancestry, the biologists will not count them as of the same species. So species is a *hybrid* relational concept, since it is individuated not only by causal powers but also by historical properties of organisms.

Consider the concept of being a planet. It is an explanatory kind in astronomy, a discipline, like evolutionary biology, the scientific status of which should be unquestioned. Like the concept of species, being a planet is a hybrid relational

concept. For whatever counts as a planet not only needs to possess a certain physical constitution but also needs to bear a certain relationship to things external to it, specifically a star that it orbits. The earth is a planet because it orbits the sun. An object that does not orbit a star will not be regarded as a planet even if it is a physical duplicate of the earth.

The identity conditions of these two concepts are *in part* determined by things external to them. In some scientific disciplines, the central explanatory kinds are even *wholly* determined by external conditions. Take the concept of money, a central concept in economics, as an example. Many causal generalizations refer to this concept. Gresham's law, for example, holds that bad money drives good money out of circulation. Money, whether good or bad, however, is *not* individuated by its physical constitution but by its causal ancestry, namely being issued by the government. In brief, money is a *pure* relational concept.

As the above examples show, many entities and properties in the sciences are taxonomized relationally, so if we agree that a plausible constraint on science must conform to actual scientific practice, then we should conclude that (GI) is false. Consequently, Fodor's argument for individualism is *unsound*.

3.2 Can Global Individualism be reconciled with the Relational Taxonomies?

Fodor, however, believes that global individualism can do justice to the prevalence of relational taxonomies in sciences. He writes:

It is patent that taxonomic categories in science are often relational. Just as you'd expect, relational properties can count taxonomically whenever they **affect causal powers**. Thus, 'being a planet' is a relational property par excellence, but it's one that individualism permits to operate in astronomical taxonomy (My emphasis) (Fodor (1987): p. 43)).

As the above passage suggests, Fodor does not think that any relational property must be ruled out by individualism. Being a planet, as I have pointed out, is plainly a relational property, since for anything to be a planet, it must stand in a certain relation to its environment, namely orbiting a star. Nonetheless, this property, according to Fodor, is one meeting the demands of individualism. For *being a planet affects the causal powers of whatever instantiates it*. So Fodor seems to think that individualism should be more *liberally* construed: *it should permit explanatory kinds to be individuated not only by narrow causal powers, but also by what affects narrow causal powers*.

But what exactly does it mean to say that being a planet (causally) affects its instances' (narrow) causal powers? Unfortunately, Fodor (1987) provides no explicit characterization of this notion. But I suggest Fodor would say something like the following. Whatever instantiates the relational property of being a planet must stand in a particular relation to a particular object (a star). And though this environmental

fact does *not constitute* the narrow correlate of the planet – its physical compositions – its presence *in a sense* plays a causal role in determining what an object does at a particular time. Or, to put it in another way, the relation that a planet stands in is *in a sense causally relevant* to its narrow behaviour. “Whether you are a planet affects your trajectory, and your trajectory determines what you can bump into” ((1987): p.43), as Fodor puts it. It is in this sense that a planet is causally affected by its being a planet and hence being a planet is a property permitted by global individualism.

Of course, not all relational properties affect their instances. Recall the **H-T** particle case discussed in chapter 2. Call a particle an “**H**-particle at time *t*” iff a particular dime of Fodor’s is heads up and a particle a “**T**-particle at time *t*” iff the dime is tails up. So Fodor can change every particle in the universe from an **H**-particle to **T**-particle and back again by turning over his dime. According to Fodor, since “whether something is an **H**-(**T**-) particle is irrelevant to its causal powers” ((1987): p.33), in the sense that the presence of the external fact (the orientation of Fodor’s dime) that figure in the individuation conditions of these two properties makes *no* causal difference to what these particles do, the relational kind **H**-(**T**) particles should be ruled out by individualism.

The above discussion suggests that we can reformulate global individualism

(premise 1) as follows:

(1*) Any explanatory categories or kinds devised by the sciences **K** must be:

(a) individuated by narrow causal powers *or* (b) individuated by what affects narrow

causal powers: the presence of the external conditions of **K** is *in a sense causally*

*relevant to its instances' narrow behaviours*⁶³.

3.3 Two Senses of Causal Relevance of External Conditions

Fodor believes that so construed, (1*) (call it *liberal* global individualism) can do

justice to the fact that in some scientific disciplines, causal properties are relationally

taxonomized. But what exactly does Fodor mean by “causal relevance”? Again, Fodor

provides no explicit answer to this question. Nonetheless, to help answer it, we may

distinguish between two senses in which the external condition(s) of a relational kind

R is (are) causally relevant to its instances' narrow behaviour.⁶⁴

(CR1): For any relational property **R**, its external condition(s), **E**, is (are) *causally*

⁶³ Stich, in characterizing his principle of autonomy, also seems to use “individualism” in this liberal sense, as after saying that “*the basic idea of the principle is that the states and processes that ought to be of concern to the psychologist are those supervene on the current, internal, physical state of the organism,*” he adds that “*historical and environmental facts will be psychologically relevant only when they influence an organism's current, internal, physical states*” ((1983): pp.164-5).

⁶⁴ This distinction between two senses of causal relevance is inspired by the one drawn by R. Stalnaker (1989; 1999: (pp.190-1)).

*relevant*₁ to the narrow behaviour of its instances: *there cannot be external conditions (e.g., environmental facts) other than E that can cause R's instances to behave as they actually do.*

(CR2): For any relational property **R**, its external condition(s), **E**, is (are) *causally relevant*₂ to the narrow behaviours of its instances: *were E not present, R's instances would not behave as they actually do.*

As is shown by the different arguments for externalism, the narrow counterpart of a given intentional property can have various causal ancestries. For example, in the twin earth story I discussed in chapter 2, even if Oscar and his replica Oscar* have the same neurophysiological state that causes them to behave in a certain way, their causal ancestries are different (Oscar's is caused by the fact that there is water (H₂O) in his environment, Oscar*'s by the fact that there is twater (XYZ) in his environment). In the case of my seeing Lorraine, the narrow counterpart of the thought that the girl standing there is Lorraine, the internal state that disposes me to exhibit certain behaviours, such as the utterance "Hi, Lorraine," surely could be produced by external conditions other than Lorraine herself – for example, by her hologram. So the external conditions of intentional properties are not causally

relevant¹ to the narrow behaviour of their instances.

Is this sense of “causal relevance” the one Fodor intends? The answer seems to be No. For as Stalnaker ((1989)) points out, like external conditions that figure in the identity conditions of folk psychological intentional states, orbiting a star is not causally relevant to the planet’s behaviour in this sense, since it is easy to conceive a possible world in which there is a large chunk of matter, which is a physical duplicate of a planet, yet instead of orbiting the sun, is in a field of force exactly like the one that the earth actually is in. So, if causal relevance is interpreted in this sense, individualism will rule out being a planet as a proper scientific kind and thus cannot be true of actual taxonomic practices in the sciences. *This means that premise (1*) is false and hence the argument still turns out to be unsound.*

It seems to me that the sense of “causal relevance” that Fodor has in mind is the second one. A planet’s behaving in a certain way at a particular time – say, having a particular velocity – is determined not only by its intrinsic features but also by the environment it is in. As a result, had the earth’s environment been changed, then, of course, the earth would have behaved in a different way. The following counterfactual seems to be true: Had the earth not been in an environment in which there is a star near it, then it would not have behaved as it actually does. And it is in this sense that the external conditions of the property of being a planet are causally relevant to the

narrow behaviour of its instances.

As Fodor **H-T** particle case illustrates, not all external conditions of relational properties are causally relevant² to their instances' behaviour. The orientation of Fodor's dime (**H** or **T**) is in *no* sense causally relevant to the particles' behaviour. Being in an environment in which the orientation of Fodor's dime is **H** or **T** does not determine how particles behave. The following counterfactuals seem to be false: (1) "Had the orientation of Fodor's dime not been heads up, then the particles would not have behaved as they actually did." (2) "Had the orientation of Fodor's dime not been tails up, then the particles would not have behaved as they actually did."

Certain external conditions of a relational property are causally relevant² to its instances' behaviour, according to Fodor, just in case there are *causal laws* that correlate the external conditions and the instances of the relational property in question. As Fodor put it "where there are no causal laws about a property, having the property – or failing to have it – has no effect on causal powers" ((1987): p. 43). In the case of the property of being a planet, such causal laws are not difficult to find. What accounts for the fact that being in an environment in which there is a star is causally relevant to a planet's behaviour is Newton's law of universal gravitation. It is in virtue of a planet's being in that environment that the star, qua a massive object, exerts a gravitational force on it and thereby the planet behaves in a certain way. By

contrast, in the case of being an H/T particle, no causal law correlating particles' behaviour with the orientation of Fodor's dime can be found.

With the notion of causal relevance², we now can flesh out global individualism as follows:

(1*) Any explanatory categories or kinds **K** devised by the sciences must be:

(a) individuated by narrow causal powers or (b) individuated by what *affects* narrow causal powers: Had the external conditions of **K** (**E**) not been present, **K**'s instances would not have behaved as they actually do. Or, equivalently, there are causal laws that correlate the external conditions of **K** with the behavior of its instances.

It is obvious that (1*) so modified, together with premises 2 and 3, suggests that we need to refine individualism in psychology (5) as follows:

(5*) Explanatory kinds devised by psychology must either be (a) autonomous properties or be (b) a kind of relational property, the external conditions of which are causally relevant ² to its instances' narrow behaviour.

It appears that Fodor believes that (5*) will still rule out the intentional properties of folk psychology as proper candidates for scientific psychology. As he says:

the property of being a planet is taxonomic because there are causal laws that things satisfy in virtue of being planets. By contrast, the property of living in a world in which there is XYZ in the puddles is *not* taxonomic because there are *no* causal laws that things satisfy in virtue of having *that* property. And similarly for the property of living in a speech community in which people use 'brisket' to refer to brisket of beef (Fodor's emphasis) ((1987): p. 43).

Is Fodor right on this point? Some think not, since wide intentional properties do causally affect the narrow causal powers of their instances – that is, the external conditions of intentional properties and their instances are related in such a way that the former is causally relevant² to the latter. (see, e.g., Stalnaker (1989); (1999): p.188, Wilson (1995): p.48)). To illustrate, consider the belief (type) that water is a boring drink. The external condition that figures in this belief's individuation condition (that of living in a world in which there is H₂O) surely is causally relevant² to the narrow or autonomous behaviour of whoever instantiates that belief type. This can be shown by running the following counterfactual test: Had Oscar not been placed in an environment in which there is water, then he would not have the neurophysiological state that serves as the narrow counterpart of that belief type and had he not had this state, he would not have behaved as he actually does. Likewise, had I not been in an environment in which the girl in front of me is Lorraine, I would not have the internal state that disposes me to instantiate certain autonomous behaviour such as waving my hand and producing the utterance "*Hi Lorraine.*"

As we saw in section 2, Fodor thinks that the project of revising wide intentional

states is pressing only because Fodor thinks that individualism in psychology conflicts with folk psychology. However, *if (GI) is liberally construed, individualism in psychology is perfectly compatible with wide intentional states*. As a result, Fodor's account of narrow content is pointless.

Moreover, even if liberal (GI) can do justice to explanatory kinds that are hybrid relational properties, it is not clear how it is compatible with the fact that the explanatory kinds of many scientific disciplines are *pure* relational properties. The concept of money, for example, is not individuated by narrow causal powers at all. Nor is it individuated by what affects causal powers.

3.4 Conclusion: The failure of the Explanan Argument

One way to make (GI) conform to the relational taxonomies commonly used in the sciences is to enlarge the notion of causal powers in premise 1.⁶⁵ If we take certain environmental and historical properties as constituents or components of causal powers, then surely (GI) is consistent with the prevalence of relational taxonomies. Suppose, for example, that we regard the evolutionary history of a particular species as one constituent of its causal powers. Then it is true that the concept of species can be described as being individuated by causal powers. For two organisms will be

⁶⁵ See Wilson (1995): pp.45-6.

classified as the same species if they share the same morphological and genetic properties and evolutionary history. However, if the notion of causal powers is interpreted in this way, Fodor's argument is still unsound. For causal powers so construed will not be determined by or supervenient on intrinsic physical properties: two individuals sharing all the same intrinsic physical properties will not necessarily agree in their causal powers in this sense, for an individual's intrinsic physical properties are insufficient to fix its environmental and historical properties. *This means that premise 2 is false and Fodor's argument turns out to be unsound. As a result, individualism cannot be deduced from the argument.*

So it seems that the only way to make both (1) and (2) true, as Wilson (1995: ch 2) points out, is to *equivocate* on the meaning of "causal powers": "Causal powers" in (1) is understood as narrow causal powers, but in (2) as wide causal powers or causal properties. This equivocation of course means Fodor's argument is *invalid*. To sum up, *Fodor's argument either contains false premises or is invalid.*

Chapter 5: Conclusion

It is generally agreed that the theoretical posits of folk psychology – intentional properties – are individuated by their propositional contents. According to externalism, this ordinary, intuitive notion of content is wide, in the sense that it is not supervenient on the intrinsic physical properties of individuals. The conjunction of these two claims – that intentional properties are individuated by propositional contents, which are wide – entails that folk psychology itself is non-individualistic.

The wideness of folk psychological contents apparently clashes with individualism, the doctrine that explanatory kinds in scientific psychology must be individualistic. As a result, if individualism can be shown to be a plausible constraint on scientific psychology, then folk psychology cannot be a legitimate part of scientific psychology. This, together with the claim that only the theoretical posits of the sciences are real, provides one route to eliminativism.

In this thesis I have been concerned mainly with the question of whether individualism is a plausible constraint on scientific psychology. But as I noted in the Introduction, all the premises of the above argument are controversial. For example, though the view that intentional states as folk psychology conceives them are wide is shared by many individualists and anti-individualists, it is by no means a generally accepted view. Some philosophers, for example, reject the semantic intuitions about

the twin earth thought experiments by arguing that contrary to semantic externalism, meanings are in the head.⁶⁶ Moreover in this thesis, following many philosophers, I have characterized folk psychology as an empirical theory containing a cluster of causal generalizations couched in terms of intentional states. This view has in fact been challenged by many philosophers and psychologists. According to simulation theory, our ability to explain and predict human behaviour by using intentional idioms is not due to our employing a theory. Rather, what enables us to do this is that we use our own decision-making systems to simulate the psychological mechanisms of others.⁶⁷ If this view turns out to be true, then it seems that the eliminativist argument cannot get off the ground, since folk psychology is not a theory after all.

I approach the question of whether scientific psychology should be individualistic by examining two distinct types of arguments. One argues for individualism from a normative claim about how the explananda of psychology should be individuated (the argument from below); the other derives individualism from a claim about how the explanantia of psychology should be typed (the argument from above). As I have shown, these two normative doctrines can be derived from two general constraints on scientific taxonomy. There are metaphysical individualism, the view that the causal powers of anything are determined by that thing's intrinsic physical properties, and

⁶⁶ See, e.g., Searle (1983).

⁶⁷ Defenders of this view include Gordon (1986) and Heil (1986).

global individualism, the view that explanatory kinds or categories in the sciences are individuated by their causal powers. In chapter 3, I argue that contrary to what Fodor believes, violating metaphysical individualism does not make a mystery of mental causation, once we recognize that there are two kinds of causal powers. And in chapter 4, I argue that global individualism is not supported by actual individuating practices in sciences, nor can it be derived from a metaphysical doctrine concerning the nature of causal explanations. All these provide us with reasons for rejecting eliminativism.

One should note that other than arguments that derive individualism from certain metaphysical and methodological doctrines, there are arguments that derive it from the considerations about the theoretical commitments and explanatory practices of a given branch of scientific psychology.⁶⁸ Burge's (1986) argument for individualism can be seen as of this type. The arguments that I have discussed – the aim of psychology argument and the explanandum argument – are of the former kind. I remain wholly silent on the latter arguments, and I admit that a complete answer to the question of whether or not scientific psychology should be individualistic cannot be settled merely by considering metaphysical principles. To evaluate the plausibility of individualism in psychology more comprehensively, one needs to examine the

⁶⁸ Classifying arguments for individualism in this way can be found in Egan (1993) and Cain (2002).

theoretical commitments of different branches of psychology. It may turn out that individualism is an adequate constraint on some but not all branches of psychology.

END

Bibliography

- Botterill, G., and Carruthers, P. (1999). *Philosophy of Psychology*. Cambridge: Cambridge University Press.
- Braun, D. (1991). "Content, Causation, and Cognitive Science". *Australasian Journal of Philosophy* 69:375-89.
- Burge, T. (1979). "Individualism and the Mental", in P. French, T. Uehling, Jr., and H. Wettstein, eds., *Midwest Studies in Philosophy, Vol. 4: Metaphysics*. Minneapolis: University of Minnesota Press.
- Burge, T. (1986). "Individualism and Psychology". *Philosophical Review* 95: 3-45.
- Burge, T. (1989). "Individuation and Causation in Psychology". *Pacific Philosophical Quarterly* 70: 303-22
- Burge, T. (1993). "Mind-Body Causation and Explanatory Practice". in Heil and Mele (1993).
- Cain, M. J. (2002). *Fodor: Language, Mind and Philosophy*. Blackwell Publishers Ltd.
- Caplan, A (1980). "Have Species Become Declasse?". *The Philosophy of Science Association*. 1980:71-82.
- Charles, D & Lennon, K (ed.). (1992). *Reduction, Explanation, and Realism*. Oxford Clarendon Press.
- Clark, A. (1993). *Associative Engines*, Cambridge, Mass.: Bradford Book. MIT Press.
- Corbi, J & Pardes, J (2000). *Minds, Causes, and Mechanicisms*. Blackwell Publishers.
- Crane, T. (1991). "All the Difference in the World", *Philosophical Quarterly* 41: 1-25.
- Crane, T. (2002). *Elements of Mind: An Introduction to The Philosophy of Mind*. Oxford University Press.

Cummins, R. (1983). *The Nature of Psychological Explanation*. Bradford Book, MIT Press, Cambridge Mass.

Cummins, R. (1991). "Methodological reflections on belief". In R. Bogdan ed., *Mind and Common Sense*, Cambridge University Press.

Davidson, D. (1963). "Actions, Reasons, and Causes", *Journal of Philosophy*, 60: 685-700.

Davies, M (1986). "Externality, Psychological Explanation, and Narrow Content". *Proceedings of the Aristotelian Society*, suppl. Vol. 60:262-83.

Egan, M. (1991). "Must Psychology Be Individualistic?", *Philosophical Review* 100:179-203.

Enc, B (1995), "The Units of Behavior". *Philosophy of Science* 62: 523-42.

Evan, G. (1982). *The Varieties of Reference*. Oxford University Press.

Fodor, J. (1974). "Special Sciences", *Synthese* 28:77-115.

Fodor, J. (1987). *Psychosemantics: The problem of Meaning in the Philosophy of Mind*. Bradford Book, MIT Press, Cambridge Mass.

Fodor, J. (1990). *A Theory of Content and Other Essays*, Bradford Book, MIT Press, Cambridge Mass.

Fodor, J. (1991). "A Modal Argument for Narrow Content". *Journal of Philosophy*, 88:5-26.

Fodor, J. (1994). *The Elm and The Expert: Mentalese and Its Semantics*. Bradford Book, MIT Press, Cambridge Mass.

Fodor, J. (1998). *Concepts: Where Cognitive Science Went Wrong*, Oxford Clarendon Press.

Gordon, R. (1986). "Folk Psychology as Simulation". *Mind and Language*, 1, 158-71.

Heal, J. (1986). "Replication and Functionalism". In J. Butterfield ed., *Language*,

Mind and Logic.

Heil, J. (1992). *The Nature of True Minds*, Cambridge University Press.

Heil, J. and Mele, A (ed) (1993). *Mental Causation*, Oxford University Press.

Horgan, T & Graham, G. (1990). "In defense of Southern Fundamentalism". *Philosophical Studies*, 62: 107-34.

Joe Lau (1993). "Three Motivations for Narrow Content".
<http://www.hku.hk/philodep/joelau/phil/NarrowContent.htm>.

Kim, J. (1989). "Mechanism, Purpose, and Explanatory Exclusion", *Philosophical Perspectives* 3 (1989): 77-108.

Kim, J. (1990). "Supervenience as a Philosophical Concept". *Metaphilosophy* 21: 1-27.

Kim, J. (1993). *Supervenience and Mind*. Cambridge: Cambridge University Press.

Kim, J. (1996). *Philosophy of Mind*. Westview Press. Cambridge: Cambridge University Press.

Lewis, D (1972). "Psychophysical and Theoretical Identifications". *Australasian Journal of Philosophy*, 50:249-58.

Lowe, E. J (2000). *An Introduction to the Philosophy of Mind*.

Macdonald, C., and G. Macdonald (eds.). *Philosophy of Psychology: Debates on Psychological Explanation*. Oxford, Blackwell.

Mayr, E. (1982). *The Growth of Biological Thought*. Cambridge, MA: Harvard University.

McCulloch, G. (1995). *The Mind and its World*. London: Routledge.

McDowell, J. (1984). "De re senses". *Philosophical Quarterly*, 34. Reprinted in C.

McDowell, J. (1986). "Singular Thought and the Extent of Inner Space". In Pettit and

J. McDowell eds., *Subject, Thought and Context*, Oxford University Press.

Wright ed., *Frege: Tradition and Influence*, Blackwell, 1984.

Mellor, D. H. (1977). "Natural Kinds". *British Journal for the Philosophy of Science* 28: 299-312.

Putnam, H. (1967). "The Nature of Mental States" In W. Capitan and D. Merrill eds., *Art, Mind and Religion*, University of Pittsburgh Press. Reprinted in his (1975).

Putnam, H. (1973). "Meaning and Reference". *The Journal of Philosophy*. 699-711. Reprinted in Moore, A. W. (eds) *Meaning and Reference*. (1993).

Putnam, H. (1975), "The Meaning of Meaning", in K. Gunderson, ed., *Language, Mind and Knowledge*. Minneapolis: University of Minnesota Press, Reprinted in H. Putnam, *Mind, Language, and Reality: Philosophical Papers, Vol. 2*. New York Cambridge University Press.

Ruse. M (ed.). (1998). *Philosophy of Biology*. Prometheus Book.

Searle, J. (1983). *Intentionality*. New York: Cambridge University Press.

Segal, C. (2000). *A Slim Book about Narrow Content*. A Bradford Book, MIT Press, Cambridge Mass.

Smart, J. C.C. (1959). "Sensations and Brain Processes". *Philosophical Review*, 68: 141-56.

Stalnaker, R. (1989). "On What is in the Head". *Philosophical Perspectives*, 3: *Philosophy of Mind and Action Theory*; reprinted in Stalnaker (1999).

Stalnaker, R. (1990). "Narrow Content", *Propositional Attitudes: The Role of Content in Logic, Language and Mind* (C. Anthony Anderson and Joseph Owens, eds.); reprinted in Stalnaker (1999).

Stalnaker, R. (1993). "Twin Earth Revisited". *Proceedings of the Aristotelian Society*, 1993; reprinted in Stalnaker (1999).

- Stalnaker, R. (1999). *Context and Content*. Oxford University Press.
- Sterelny, K. (1990), *The Representational Theory of Mind*. Blackwell Publishers.
- Stich, S.(1978), "Autonomous Psychology and Belief-Desire Thesis", *Monist* 61:573-91.
- Stich, S. (1983). *From Folk Psychology to Cognitive Science: The Case Against Belief*. Bradford Book, MIT Press, Cambridge Mass.
- Stich, S. (1996). *Deconstructing the Mind*. Oxford University Press.
- Von Eckardt, G. (1993). *What Is Cognitive Science?* Cambridge, Mass.: Bradford Books/MIT press.
- Walsh, D. M. (1999). "Alternative Individualism". *Philosophy of Science*, 55: 628-48.
- Wilkes, K. (1991). "The Relationship between Scientific Psychology and Commonsense Psychology". *Synthese*, 89 15-39
- Wilson, R. (1995). *Cartesian Psychology and Physicalism*, Cambridge University Press.
- Wilson, R. (1999). *Species: New Interdisciplinary Essays*. A Bradford Book, MIT Press, Cambridge Mass.
- Woodfield, A. (ed.). (1982). *Thought and Object*. Clarendon Press, Oxford.

Acknowledgement

I should like to express my gratitude to my supervisor, Professor Wong Kai Yee, from whom I have learnt much throughout the past two years in the department of Philosophy. I would like to thank also Professor Christopher J. Fraser. Not only did he provide me with many valuable comments, he also helped me polish my English writing. Finally, my warmest thanks are reserved for Lorraine. Without her whole-hearted support and encouragement, this thesis would not have been completed on time.

CUHK Libraries



004078284