Physicalism, Mind, and the Ontology of Properties

KAIDA, DAISUKE

Physicalism, Mind, and the Ontology of Properties

Daisuke Kaida

Department of Philosophy

University of Durham

Abstract

In this thesis, I will try to find out how we can accommodate human mind in a physical world. What is significant in my attempt is that I try to approach this problem with a prospect that the theory of mental causation and consciousness will be subsumed under *the general theory of properties*.

With this general orientation, I will do the following three things in this thesis:

(1) I will trace and develop the theories of mental causation and consciousness that have been conducted in the philosophy of mind.

(2) I will trace and develop the general theories of properties that have been conducted in analytic metaphysics.

(3) I will try to show that there is a prospect of combining the theory of mental causation and consciousness with the general theory of properties (or rather subsume the former under the latter).

In Chapter 2, I will clarify what physicalism should be. I formulate a minimal version of physicalism. The reason why I take physicalism is also stated. In Chapter 3, I survey the history of non-reductive physicalism, and examine a version of non-reductive physicalism, Anomalous Monism. I will show that Anomalous Monism cannot explain the causal efficacy of mental properties. In Chapter 4, I will examine Jaegwon Kim's attack on non-reductive physicalism – the causal exclusion problem. I will survey several defences from non-reductive physicalist camp and show that none of them is successful. In Chapter 5, I will examine Kim's solution to the causal exclusion problem. After I defend Kim's position from possible objections, I will make clear some consequences of Kim's position. These chapters are devoted to setting the background for discussing mental causation.

Chapter 6, 7, 8, and 9 compose a main part of the thesis. Chapter 6 and Chapter 7 are devoted to the ontology of properties. In these chapters, I will try to formulate and defend a Causal Trope Theory of Properties, which is a causal theory of properties combined with a trope theory. Chapter 6 starts with the examination of John Heil's view on the ontology of properties and objects. I will then expound my own view on properties and compare it with Heil's view. After that, I will try to defend the Causal Trope Theory against a rival theory, Humean Theory. In Chapter 7, I will defend my

own view (a version of a dispositionalist view) from a typical and influential categoricalist view on properties – Prior, Pargetter, and Jackson's view. I will also attack the categoricalist view.

After I set the background for the mental causation problem and the general theory of properties, I will go on to apply these results to the mental causation problem. In Chapter 8, I will show how David Robb's argument brings in new aspects of the problem. I will show that Robb's view has a similar consequence as Kim's but it gives us a more precise ontological picture. In Chapter 9, I will examine Shoemaker's view on mental causation. I will show that Shoemaker's view is the most prospective and successful option so far.

In Chapter 10, I will survey the main theories of consciousness (qualia) and intentionality. After I explain why consciousness (or qualia) is a serious problem for physicalists, I try to defend the representation theory of consciousness. I will show that if we can explain consciousness by intentionality, and explain intentionality in a physicalistic framework, then we can understand consciousness in a physicalistic framework. I also focus on the normativity character of intentionality and try to defend a teleological approach to intentionality.

Finally, in Chapter 11, I will show how the problem of consciousness (qualia) could be viewed as some versions of the problem concerning the general theory of properties. I will show that which view in the general theory of properties we take has some effects on which view in the theory of qualia we should take. It will be shown that the view about the theory of properties which I take in Chapter 6 and Chapter 7 and the view in the theory of consciousness (the representation theory of consciousness) which I take in Chapter 10 are good combination, and that I have good reason to support the representation theory of consciousness discussed in Chapter 10.

In the appendix, I will, quite briefly, consider another problem of mental causation, the problem of the extrinsicness of mental properties. I will show why this problem is not a problem for physicalism alone and suggest a possible way to solve it.

CONTENTS

1 Introduction

In this thesis, I will try to find out how we can accommodate human mind in a physical world. On one hand, many of us believe that scientific theories, especially the ultimate physical theory, get at the truth of the matter more or less. We believe that our world consists of physical objects and properties and that the physical objects and properties behave as the physical theory insists, more or less. This is an intuition of *physicalism*. On the other hand, we also believe that we have our *minds*. Suppose William notices Mary who is on the other side of a road. He wants her to notice him, he believes that if he raises his hand she will notice him, and he raises his hand. In this situation, he has a series of mental experiences: he feels a conscious feeling of seeing a colour; he has a desire and a belief; he also has a belief that his desire and belief cause the rising of his hand. All these events (or states) are mental phenomena. Now there is a problem. Are these two world views compatible with each other? It seems that the physicalistic world view does not include mental phenomena. It seems, moreover, that the physicalistic world view excludes mental phenomena. We, therefore, need to find out a way to accommodate mental phenomena in a physical world. To be more specific, what I will investigate in this thesis is the problem of *mental causation* and the problem of *consciousness*. I will try to explain these mental phenomena within a physicalistic world view.

This is not a new problem, of course. The problem of accommodating the human mind in a physical world is a version of the traditional mind-body problem. What is significant in my attempt, however, is that I try to approach this problem with a prospect that the theory of mental causation and consciousness will be subsumed into *the general theory of properties*.

With this general orientation, I will do three things in this thesis:

(1) I will trace and develop the theories of mental causation and consciousness that have been conducted in the philosophy of mind.

(2) I will trace and develop the general theories of properties that have been conducted in analytic metaphysics.

(3) I will try to show that there is a prospect of combining the theory of mental causation and consciousness with the general theory of properties (or rather subsume the former into the latter).

In Chapter 2, I will clarify what physicalism should be. I formulate a minimal version of physicalism. The reason why I take physicalism is also stated. In Chapter 3, I survey the history of non-reductive physicalism, and examine a version of non-reductive physicalism, Anomalous Monism. I will show that Anomalous Monism cannot explain the causal efficacy of mental properties. In Chapter 4, I will examine Jaegwon Kim's attack on non-reductive physicalism – the causal exclusion problem. I will examine several defences from the non-reductive physicalist camp and show that none of them is successful. In Chapter 5, I will examine Kim's solution to the causal exclusion problem. After defending Kim's position from possible objections, I will make clear some consequences of Kim's position. These chapters are devoted to setting the background for arguing mental causation.

Chapter 6 and Chapter 7 are devoted to the ontology of properties. In these chapters, I will try to formulate and defend a Causal Trope Theory of Properties, which is a causal theory of properties combined with a trope theory. Chapter 6 starts with the examination of John Heil's view on the ontology of properties and objects. I will then expound my own view on properties and compare it with Heil's view. After that, I will try to defend the Causal Trope Theory against a rival theory, Humean Theory. In Chapter 7, I will defend my own view (a version of a dispositionalist view) from a typical and influential categoricalist view on properties – Prior, Pargetter, and Jackson's view. I will also attack the categoricalist view.

After I set the background for the mental causation problem and the general theory of properties, I will go on to apply these results to the metal causation problem. In Chapter 8, I will show how David Robb's argument brings in new aspects of the problem. I will show that Robb's view has a similar consequence as Kim's but it gives us a more precise ontological picture. In Chapter 9, I will examine Shoemaker's view on mental causation. I will show that Shoemaker's view is the most prospective and successful option so far.

In Chapter 10, I will trace and develop the theories of consciousness (qualia) and intentionality. After I explain why consciousness (or qualia) is a serious problem for

physicalists, I try to defend a representation theory of consciousness. I will show that if we can explain consciousness by intentionality, and explain intentionality in a physicalistic framework, then we can understand consciousness in a physicalistic framework. I also focus on the normativity character of intentionality and try to defend a teleological approach to intentionality.

Finally, in Chapter 11, I will show how the problem of consciousness (qualia) could be viewed as some versions of the problem concerning the general theory of properties. I will show that which view in the general theory of properties we should take has some effects on which view in the theory of qualia we should take. It will be shown that the view in the theory of properties which I take in Chapter 6 and Chapter 7 and the view in the theory of consciousness (the representation theory of consciousness) which I take in Chapter 10 are good combination, and that I have good reason to develop the representation theory of consciousness in Chapter 10.

In the appendix, I will, quite briefly, consider another problem of mental causation, the problem of extrinsicness. I will show why this problem is not a problem for physicalism alone and suggest that in what direction the problem could be solved.

2 What Physicalism Should Be: A Minimal Physicalism as a Research Programme

As was stated in the introduction, the purpose of the whole thesis is to understand how our mind is realized in a physical world. To restate the problem in another way, I try to investigate how *physicalists* can accommodate mental phenomena in their world view. The first thing I must do is to make it clear what physicalism is. Of course, this is not an easy task. I can not give the answer to this difficult problem once and for all, at this stage. In this chapter, I am going to describe a rough idea of what physicalism should be. I will try to formulate a form of physicalism – *a minimal physicalism*. This will set the background on which the whole research after Chapter 3 is organized.

2 - 1 What is Physicalism?

Physicalists accord special privilege to physics. Roughly speaking, physicalists claim that only physical things (or *material* things – I don't distinguish them throughout this thesis) exist in our world. Physicalism insists that everything in this world is *nothing over and above* material things, such as the atoms, the ultimate particles, or whatever the current or future physics says there are. The size of the things doesn't matter at all: there exist ordinary macroscopic things such as tables, chairs, and pencils; there exist much smaller microscopic things such as molecules or ultimate particles; there exist much bigger things such as planets or galaxies. If physicalism is true, however, all these things are composed of microscopic particles which are posited by the physical theory; they do not have as their components non-physical things, i.e. entities which do not appear in the physical theory. We believe or desire something; we think about, remember, and image something; we perceive something and feel various sensations. These are usually regarded as *mental* activities. If physicalism is right, however, these mental activities are not conducted by some non-material (non-physical), mental objects. These purported mental activities must be explained by the physical.

It is, thus, fairly easy to grasp a rough idea of physicalism, and many scientists

and philosophers seem to share this physicalistic intuition. However, it is not so easy to provide an exact formulation of physicalism. First of all, what exactly does 'the physical' mean? And what exactly does 'the non-physical' or 'the mental' mean? What exactly is the criterion to distinguish between the physical and the mental? These are, no doubt, important questions. I think, however, that we can bypass these difficult questions for the time being, and just rely on the ordinary concepts of the physical. The physical includes physical substances such as ultimate particles, tables, galaxies and so on. The physical also includes physical properties such as charge, mass, shape and so on. The physical is the ontological item that current and future physics could capture properly. The reason why I think we can bypass these questions is because we are going to shape *a minimal physicalism* in the following. I will return to this problem later in the thesis. For the time being, I freely mention physical substances and physical properties ('the physical') without defining the meaning of 'the physical' clearly.

Setting aside the problems concerning the meaning of 'the physical' and 'the mental', there is still a problem about the *formulation* of physicalism. Although several formulations of physicalism have been suggested[1], there seem to be no agreement about *the* formulation of physicalism. I do not, however, attempt to provide *the* formulation of physicalism in this thesis. That is to say, I do not try to give the *necessary and sufficient* conditions for physicalism. What I will try to do in the following is to give three *necessary conditions* for physicalism. This is to formulate *a minimal* physicalism. As I see it, physicalism requires, at least, three principles: the principle of the denial of non-physical objects, the principle of physical determination, and the principle of physical causal closure. Let us see them one by one.

2 - 2 Requirement One: The Principle of the Denial of Non-physical Objects

The first requirement for physicalism is the denial of non-physical objects. It can be formulated as follows:

> **The Principle of the Denial of Non-physical Objects (DNO)**: Every object in the world is a physical object.

[1] See Chapter 2 of Poland (1994) for a survey of several formulations of physicalism.

The main point of this requirement should be clear. This requirement tries to exclude, first of all, non-physical *substances* – the Cartesian soul, for instance – from a physical world. Therefore, the defender of DNO clearly commits oneself to Anti-Cartesian. This commitment is appropriate. If one is a physicalist, one cannot believe in Cartesian soul.

However, does DNO deny only the non-physical substances – the Cartesian soul? Does DNO deny non-physical *properties* as well? That depends on how we interpret 'the non-physical objects' in the formulation. If we include *properties* in the category of 'object' in DNO, then DNO clearly denies the existence of non-physical properties, as well as non-physical substances. However, I do not take this way. That is because I am trying to formulate a *minimal* physicalism. It is an open question if physicalism should deny non-physical properties. If one accepts the existence of mental substances, one cannot be a physicalist at the same time. However, many philosophers believe that one can both accept the existence of mental properties and be a physicalist at the same time. In fact, as I will discuss it in the next chapter, the dominant view in the philosophy of mind since the late 1960th has been a version of the latter type of physicalism – non-reductive physicalism. Non-reductive physicalism has it that while there exist only physical substances in this world, some of the physical substances, when composed in some special ways, have mental properties as well that cannot be reduced to physical properties. If one formulates physicalism as the denial of any non-physical items, one cannot take any kind of non-reductive physicalism: one has to take either reductive physicalism or eliminativism. Even if non-reductive physicalism were to be denied, it should be denied as a consequence of arguments. We should not exclude the possibility of such position right from the beginning.

I will, therefore, interpret 'object' in DNO as the ontological items that are particularized and are bearers of properties. Substances are objects in this sense. And I interpret 'physical object' as objects which possess at least one physical property. Therefore, DNO insists that there cannot be an object which has *only* non-physical properties. The typical example of such objects is, off course, the souls that Descartes believed in. Notice again that DNO does not deny the existence of mental properties. It insists that if an object has a mental property, it must have physical properties as well.

Now I will mention further qualification concerning objects. In the category of

objects, I include substances. What about events, processes, and states? Whether we can include events in the category of 'object' in DNO depends on the definition of events. According to Davidson's view ('coarse grained' view) on events, one and the same event can possess both mental properties and physical properties[2]. Therefore, we can state, as a minimal requirement for physicalism, that every event in the world is a physical event; this statement does not preclude the possibility that an event might have mental properties as well. However, according to Kim's view ('fine grained' view) on events, where events are defined as property exemplifications, one and the same event cannot have both mental properties and physical properties at the same time: they must be different events with each other[3]. Therefore, if we state that every event in the world is a physical event, this statement precludes the possibility that there exist mental properties. This preclusion is not appropriate if we are looking for a minimum requirement for physicalism. In sum, if we take Davidson's view on events, we can include 'events', in the category of 'objects' in DNO; if we take Kim's view, we cannot. The same consideration applies to the concept of 'processes' and 'states' as well. If a process (state) can have different properties at the same time, then a physicalist can insist that every process (state) in the world is a physical process (state) – a process (state) with at least one physical property, without precluding the possibility that the process (state) might possess mental properties as well. A brain, for example, might be in a mental state that is not reduced to physical states. We should not preclude this possibility, as we are stating minimum requirements for physicalism.

There is one more point which should be noticed concerning this principle. Although the principle of the denial of non-physical objects captures the core of physicalistic intuition, we should notice that this principle only states the ontological claim concerning just one possible world – actual world (that is, *our* world). It only states that there is no non-physical object in our world. It does not preclude the possibility that there exists a possible world where there are immaterial, non-physical Cartesian souls. I will return to this problem in the next section where I discuss the second requirement for physicalism.

---

[2] See Davidson (1969).

[3] See Jaegwon Kim, 'Events as Property Exemplifications,' reprinted in his *Supervenience and Mind* (Cambridge: Cambridge University Press, 1993).

2 - 3 Requirement Two: The Principle of Physical Determination

An intuition of physicalism includes the claim that everything in this world is material, and the first requirement captures this aspect of the physicalist intuition. There are, however, other aspects of the physicalist intuition. Some physicalist formulate their views as insisting that the way how the world is, concerning its material constitution, *determines* the way the world is *simpliciter*.[4] To use a theological metaphor, once God creates all the physical substances and specifies their properties and relations, God does not have to add something else any more.

Frank Jackson formulates physicalism along this line. According to Frank Jackson's formulation, physicalism insists that any world which is a physical duplicate of our world is a duplicate *simpliciter* of our world.[5] To use the concept of supervenience, it insists that every property in our world supervenes on micro physical properties and relations – the properties of fundamental particles and the relations between them. Along this line, I will formulate the second requirement as follows:

> **The principle of Physical Determination (PD)**: It is nomologically necessary that two objects that share all the physical properties must share all the mental properties as well.

This requirement insists that mental properties supervene on physical properties.

Let me add some comments on the formulation. First, the supervenience in the formulation must be *strong* supervenience. *Strong* supervenience requires that two objects that share all the physical properties (we can call them 'physical duplicates') share all the mental properties as well, even if those objects inhabit in *different* possible worlds (so long as those possible worlds share all the natural laws). *Weak* supervenience, on the other hand, requires that two physical duplicates share all the mental properties only when those objects inhabit in the same world. It should be fairly clear that we need strong supervenience here. For an ontological view is not worthy of physicalism if it admits that two physical duplicates differ in their mental properties in different possible

---

[4] See, for example, Lewis (1983) and Jackson (1998).
[5] See Jackson (1998), p.12.

worlds, all the physical laws being shared by those worlds.

Secondly, we must notice that this requirement concerns properties *across* possible worlds. While the first requirement, DNO, is a claim concerning a single world, the second requirement, the principle of physical determination, is a claim *across* possible worlds. Remember that DNO concerns only one world: it denies mental substances (substances with mental properties only) in the actual world. DNO allows the possibility that a substance a with a physical property P_1 and a mental property M_1 in the actual world W_{actual} has the same physical property P_1 and a different mental property M_2 in a possible world W_2, as long as the substance a has some physical properties. In this situation, the physical property fails to *determine*, or *fix,* the mental properties. What PD denies is this situation. Once God creates the world where the substance a has the physical property P_1, God does not have to specify further whether a has M_1 or M_2 in a possible world. It just has the same mental property M_1 as in the actual world, according to PD.

Thirdly, we must also notice that supervenience does not say much about ontology. As Kim persuasively points it out, supervenience (strong or weak) does not specify ontological matters very much. What any supervenience thesis says is just that there is a *covariance* between two sets of properties. As Kim sees it, we need to add something to supervenience thesis, if we want to make our view on the mind-body problem more substantive in an ontological sense.[6] I agree with Kim in that supervenience thesis does not say much about ontology. However, because of this austereness, as I see it, the concept of supervenience is quite proper for stating a *minimal* requirement for physicalism. One can add some ontological content to the supervenience thesis if one wishes. It suffices, for the time being, to notice that the supervenience thesis works as a necessary condition for a minimal physicalism. I will take 'realization' as the ontological ground for supervenience, later.

The first requirement, DNO, is the claim which states an ontological qualification within one (actual) world. The second requirement, PD, widens the scope, and states an ontological qualification on properties across possible worlds. However, there is still a limitation about the second requirement: it does not concern the properties across *time*. In PD, we consider the possible situations with time being fixed. Now we

[6] See Kim (1998), pp. 9-15.

are going to see the third requirement, where we consider properties across time.

2-4 Requirement Three: The Principle of Physical Causal Closure

There is one more aspect of physicalist intuition: physicalism should insist that the world is physically and causally closed. If a glass on the table breaks, then there must be a physical cause of the breaking if there is any cause at all.[7] The third requirement, therefore, concerns the causal network where physical events are connected with each other across time. Suppose that a glass on the table moves at some time. If we see this event (the moving of the glass), we expect there is some physical cause of this event: someone might have pushed the glass; someone might have tilted the table and the glass might have been attracted by the gravitation force downward. If we cannot find any physical cause and if we have to accept that a non-physical cause contributes to the moving of the glass, this means that we must give up physicalism. Thus, I will formulate the third requirement for physicalism as follows:

> **The Principle of Physical Causal Closure (PCC)**: For all physical objects, if it has a cause at *t*, it has a sufficient physical cause at *t*.

Many philosophers have suggested various formulations of something like the principle of physical closure[8], and I will discuss one of them later. Here, I give some comments on the requirement in this formulation.

First, we must admit the possibility that a physical event has no cause whatsoever. According to the big bang theory, the very first event in the universe – Big Bang – does not have any cause (physical or non-physical). It just happened. To admit the existence of Big Bang is no more denying physicalism. This is why 'if it has a cause at *t*' is inserted in the requirement.

Second, if we don't mention the *time* when a cause exists, the requirement

---

[7] See, for example, Papineau (1990) and Kim (1998).

[8] Douglas Ehring, for example, formulates the principle he calls 'the completeness of physical property causation' in the following way: for every physical effect *e* that has a cause at *t*, there are physical properties $P_1 \ldots P_n$ that are instantiated at *t* that are causally sufficient for *e*. It is obviously very close to my formulation. My formulation can be understood as a simplified version of Ehring's formulation. See Ehring (2003), p362.

would be too weak. Suppose that there exist a physical event P_3 at t_3, a mental event M at t_2, and another physical event P_1 at t_1 ($t_1<t_2<t_3$). Suppose also that P_1 causes M, M causes P_3, and there is not a cause of P_3 at t_2 other than M. This case means that if we trace back a physical event, we must get out of the physical domain, which should not be acceptable for physicalism. If we don't mention the time t in the requirement, we cannot exclude this case.[9]

Third, notice that we must mention the sufficiency of the physical cause at t in the requirement. Even if a physical event P has a physical cause P' at t, if the physical cause P' is not sufficient for bringing about P, then we need an extra non-physical cause M at t to cause P together with P'. This is obviously the violation of physicalism. That is why the sufficiency of the cause is mentioned in the requirement.

At this point, someone might insist that the requirement of sufficiency of physical causes is too strong, considering what the current quantum physics tells us. According to quantum mechanics, the world is ineliminably probabilistic. Even if we are given the complete physical information about the world at a time, we cannot predict the future from the information deterministically. Therefore, the physical cause cannot be sufficient in the sense that it brings about the effect as a matter of nomological necessity, can it? I don't object the ineliminably probabilistic character of quantum mechanics and its implication about the ontology. However, I don't think I must alter the principle in the face of the indeterministic implication of quantum mechanics. The ineliminably probabilistic nature of physical processes and the sufficiency of the physical causes are perfectly compatible, if we understand the sufficiency in the sense that the world does not need any non-physical causes for the physical effects to occur. 'Sufficient physical cause' in the requirement does not mean that the physical cause necessarily (in nomological sense) brings about the effect. It might be that the effect does not occur although its purported physical cause occurs. As long as the probability of the occurrence of the effect is completely fixed by the physical cause, it is OK. The requirement just excludes the case that a non-physical cause *changes* the probability of the effect's occurrence specified by the purported physical cause: in this case, the physical closure is, of course, violated.

---

[9] This is pointed out by Lowe (2000).

2-5 A Minimal Physicalism as a Research Program

As stated above, these three requirements are offered as a minimum requirement for physicalism. Denying any one of them would lead to the denial of physicalism. However, it should be noticed that these requirements would not be the sufficient condition for physicalism. As I stated in the introduction, I will devote quite many amount of this thesis to the mental causation problem and the consciousness problem with the supposition of physicalism. That is to say, I will try to clarify whether we can understand mental causation and consciousness within the framework of physicalism or not.

But there might be a question: why start with physicalism? Can we really presuppose physicalism? Is it so sure that physicalism is, or turn out to be, true? I am not sure if physicalism is, or turn out to be, true. I am not sure how we can prove the truth of physicalism, either. I believe, however, that physicalism is worthy of pursuing as a *research program* in ontology. When we have several research programs, we can assess the prospect of the programs in some respects. What research programs do we have with us now? Roughly speaking, there are three programs: physicalism, dualism, and mentalism. All three of them have both advantages and disadvantages of their own. But, as I see it, the disadvantage of mentalism – its counter intuitiveness – is too conspicuous for us to pursue it today: it is difficult for us to really believe that every object in the world is mental. What about dualism? Dualism has long history since Descartes, and many ontological attempts have been made in this tenet. I cannot fully discuss the disadvantages of dualism here, but I can at least point out one problem. If we try to understand mental causation within the framework of dualism, we are naturally led to accept that a purely mental event without any physical energy at all could cause a physical event, such as bodily movements, with some physical energy. Trying to avoid this consequence, many dualists are led to the views such as pre-established harmony or epiphenomenalism which deny our common sense about the mental causation, our common sense that there exists the mental causation.

Also, physicalism has an advantage: its simplicity. Physicalism conducts the investigation of the world, insisting just one working hypothesis that the world is made up of the physical materials and the physical principles alone. If the physicalist

investigation is successfully accomplished in this way, our understanding of the world would progress very much. Even if it ends up in failure, it would be fairly easy, due to its simplicity, to see in what way it fails, which would help us to revise the strategy of the investigation. In sum, I believe that when two or more research programs are available and we are not very sure which one of them should be true, we should take the simplest one. This is why I take physicalism as a research program.[10]

[10] The idea of 'a minimal physicalism as a research program' comes from Mino (2004), although his formulation of the minimal physicalism is slightly different from mine. The general idea of physicalism as a research program seem to be widely shared by many physicalists, such as J. Kim, J. Fodor and others.

3 The Rise of Non-reductive Physicalism

3 - 1 A Brief History until the Rise of Non-reductive Physicalism

Many of the current philosophers of mind adopt physicalism as a research program. In this section, I will follow the development of the current philosophy of mind and show how the problem of mental causation arises in the framework of the physicalism.

The direct ancestor of the current philosophy of mind is 'the mind-brain identity theory' which arose around 1950th.[11] It insists that mental states *are* the states of cranial nerves, or, to be precise, that mental states will be identified with the states of cranial nerves by future brain neurophysiology. An experience of pain, for example, is identified with a neurophysiological state. This is not to say that one and the same object has both a mental property of being pain and a neurophysiological property, but it is to say that a property (or type) of being pain is identical with a neurophysiological property (or type). Mind-brain identity theory (or 'type identity theory') is a version of physicalism, for neurophysiological properties are naturally considered to be physical properties. If mind-brain identity theory is right, then we have at least one explanation of mental causation: mental properties are nothing but physical properties and mental causation is nothing but physical causation.

Mind-brain identity theory, however, declined very quickly. The reason of the decline is because mind-brain identity theory was attacked by the objections from two directions: the objection based on the multiple realizability of mental properties and the objection based on the anomalousness of the mental. Let us see these objections in order.

---

[11] I have in mind the works by U. T. Place, H. Feigl, and J. J. C. Smart etc. See, for example, Place, 'Is consciousness a brain process?', *British Journal of Psychology* 47, 1956; Feigl, 'The mental and the physical', in H. Feigl, G. Maxwell, and M. Scriven, eds., *Minnesota Studies in the Philosophy of Science*, Vol.2, Minneapolis, 1958: University of Minnesota Press; Smart, 'Sensations and Brain Process', *Philosophical Review* 68, 1959.

３－２ Objection from the multiple realizability

One objection to mind-brain identity theory, which is first raise by Hilary Putnam[12], is based on the multiple realizability of mental properties. This objection has it that a mental property can be realized by various physical properties. A pain, for example, can be realized in many ways. In human beings, a neural state, C-fibre excitation, realizes pain. An animal with quite different neural constitutions from human beings could feel pain, and the neural property that realizes the pain in that animal would be quite different from those of human beings. If a Martian or a robot could feel pain, their physical states that realize their pain would be quite different from those of human beings. Therefore, we cannot identify a mental property with any one of the physical properties that realize the mental property.

One might attempt, here, to identify a mental property with the disjunction of the physical properties that realize the mental property. However, it is not possible. For a disjunctive property cannot generally be regarded as a genuine property. It is difficult to regard a disjunctive property of 'being C-fibre excitation *or* being such and such Martian brain state *or* being such and such electric state' as one genuine property (I will discuss the genuineness of properties later, especially in Chapter 5, 8, and 9).

３－３ Objection from the anomalousness of the mental

There is another objection to mind-brain identity theory, which is raised by Donald Davidson independently of the objection from the multiple realizability.[13] This objection is based on the thesis that the mental does not accord with laws (i.e. the mental is anomalous). As I take up this objection later in more detail, I just sketch the outline. As Davidson insists it, while it is the principle of *causation* that governs the network of the physical, it is the principle of *rationality* that governs the network of the

---

[12] See H. Putnam, 'Psychological Predicates', in W. H. Capitan and D. D. Merrill, eds., *Art, Mind, and Religion*, 1967, University of Pittsburgh Press. Reprinted under the title of 'The Nature of Mental States' in his *Mind, Language, and Reality: Philosophical Papers, vol.2*, 1975, Cambridge University Press.

[13] See D. Davidson, 'Mental Events', in L. Foster and J. Swanson, eds., *Experience and Theory*, 1970, Duckworth. Reprinted in his *Essays on Actions and Events*, 1980, Oxford University Press.

mental. We can apply the concept of rationality to the mental, but not to the physical, whereas we can apply the concept of causation to the physical, but not to the mental. If this is right, there cannot obtain any strict law connecting the mental, neither can obtain any strict law connecting the mental with the physical. If mental type is identical with physical type, there should be a kind of law connecting the former to the latter. Therefore, we cannot identify a mental property with a physical property. I will get back to this objection later.

3-4 The Rise of Non-reductive physicalism

With the objections above, a new physicalism arose around the late 1960's. The new physicalism insists that the mental properties are not identical with the physical properties. It also insists that the mental properties are not reducible to the physical properties. In this reason, the new physicalism is called 'non-reductive physicalism'. Within non-reductive physicalism, we can distinguish two views (or two groups of views) according to the objection they make to mind-brain identity theory: functionalism and anomalous monism (Davidson's view)[14]. While the former arises from the objection based on the multiple realizability, the latter arises from the objection based on the anomalousness of the mental.

Non-reductive physicalism denies mind-brain identity theory (type identity theory) and accepts the existence of mental properties which are different from physical properties. Why can we, then, regard non-reductive physicalism as a version of 'physicalism' nonetheless? First, although it admits real mental properties that cannot be reduced to physical properties, it denies the existence of mental *substances*. Therefore, it satisfies the first requirement, the principle of the denial of non-physical objects, in section 2-2. Second, it satisfies the second requirement, the principle of supervenience of the mental properties on the physical properties. Therefore, we can say that non-reductive physicalism satisfies at least two of the necessary conditions for physicalism. As we will see shortly, however, there is a doubt that non-reductive

[14] The leading functionalists include Putnam and Fodor. See, for example, their studies in N. Block, ed., *Reading in the Philosophy of Psychology, vol.1*, 1980, Harvard University Press. I will discuss Davidson's anomalous monism shortly.

physicalism might not satisfy the third requirement for physicalism.

Non-reductive physicalism was considered to accord the multiple realizability of mental properties, for the principle of supervenience just requires that if a physical property of an object is determined, then a mental property of the object is also determined, but not vice versa. Also, non-reductive physicalism insists that the irreducibility of mental properties, which was thought to certify the autonomy of psychology along with other special sciences. Furthermore, non-reductive physicalism was thought (or expected) to explain mental causation in the following way. The principle of supervenience has it that every mental event is also physical event, which would make it possible to think that mental causation is a kind of physical causation, and mental properties have causal efficacy or causal relevance of their own not reducible to physical properties. However, this expectation got into difficulties and I will discuss this problem in chapter 4.

Let us sum up the non-reductive physicalist assertion clarified so far. It insists, first of all, the three requirements formulated in chapter 2:

The Principle of the Denial of Non-physical Objects (DNO): Every object in the world is a physical object.

The Principle of Physical Determination: It is nomologically necessary that two objects that share all the physical properties must share all the mental properties as well.

The Principle of Physical Causal Closure: For all physical objects, if it has a cause at *t*, it has a sufficient physical cause at *t*.

In addition to these, it also insists the following two assertions:

The Principle of Difference: Mental properties are different from and irreducible to physical properties.

The Principle of Causal Efficacy: Mental properties have causal efficacy or causal relevance as well as physical properties.

In the next section, I will examine a version of non-reductive physicalism, Anomalous

Monism.

３－５ Anomalous Monism and the Causal Efficacy of the Mental

In order to properly assess the current arguments on mental causation, we need to go back to Donald Davidson's Anomalous Monism and a serious problem raised against it. Anomalous Monism is Davidson's solution to the mind-body problem. In this section, I will, first, review Anomalous Monism, and then examine an objection raised against it.

３－５－１ Anomalous Monism

Anomalous Monism is the following view:

> **Anomalous Monism (AM)**: Mental events are identical with physical events (i.e. monism) and mental descriptions of events cannot conform to strict laws (i.e. anomalism).

Davidson draws AM from the following three principles, each of which seems plausible in itself:

> (P1) Mental events cause, and are caused by, physical events. (The Principle of the Reality of Mental Causation)
> (P2) If a causal relation obtains between two events, then there must be a 'strict' law connecting the descriptions (or the properties) of the cause event and the effect event. (The Principle of the Nomological Character of Causation)
> (P3) There are no 'strict' psycho-physical laws. (The Principle of the Anomalousness of the Mental)

Although it appears to be difficult to hold three principles at the same time, this apparent incompatibility is, as Davidson sees it, deceptive. Davidson insists that we can hold the three principles at the same time by adopting a version of monism about events, and he insists moreover that the version of monism about events is drawn from these

principles as a logical consequence. The argument goes like this. Suppose a mental event *c* caused a physical event *e* (P1). By (P2), there must exist a strict law connecting a description (property) of *c* and a description (property) of *e*. However, by (P3), a mental description (i.e. a description by mental predicates) cannot appear in a strict law. Therefore, there is no possibility other than the mental even *c* has a physical description (property) as well as the mental description (property) and *that* physical description (property) appears in the strict law. It is, thus, concluded that every mental event has some physical descriptions (properties) as well, and, therefore, is a physical event as well. Therefore, it is not that there are two kinds of events, mental events and physical events, in the world, but that there is just one kind of neutral event, describable both by mental predicates and physical predicates. When we call Davidson's view on mind-body problem 'Anomalous Monism', it might refer to just [AM], or [AM] + principles (P1)-(P3).[15]

Let me add a few brief comments. First, in the argument outlined above, a description and a property of events are not distinguished. We must notice that Davidson does not commit himself to the existence of properties in the sense that properties are distinguished from descriptions. However, in the arguments on mental causation starting from Anomalous Monism, many philosophers have not distinguished clearly between properties and descriptions. Therefore, I do not distinguish between them for the time being. I will consider this point later.

Second, it is very important to ask what is the ontological status of causal relata. As a candidate for causal relata, one could think of various ontological items such events, facts, and properties. Even if one chooses events as causal relata, there still remains a problem of what kind of ontological theory of events one would take. As Davidson himself sees it, a causal relatum is an event as a particular, a particular that could be described in many ways and that in itself does not have any essential character at all. The problem of the ontological status of causal relata will be argued in detail later.

3-5-2 Anomalous Monism as a Non-reductive Physicalism

[15] See Donald Davidson, 1970, 'Mental Events'.

Anomalous Monism is plausibly considered to be a version of non-reductive physicalism. It insists that every mental event is also a physical event (and not vice versa). This means that there cannot be a *purely* mental event, an event that has only mental properties (lacks physical properties), which satisfies the Principle of the Denial of Non-physical Objects. It also insists on the anomalousness of the mental, which says that there cannot be any strict law connecting mental properties with other mental properties (a strict mental-mental law), nor can there be any strict law connecting mental properties with other physical properties (a strict mental-physical law). According to Nagel's conception of reduction, a reduction of a theory to another theory is conducted by deriving the laws of the former from the laws of the latter. And, for the derivation, we need 'bridge laws', laws connecting the basic predicates (concepts) of the former with those of the latter[16]. If there cannot be any strict mental-physical law, then we cannot have the bridge laws which are needed to reduce mental properties to physical properties.

Denying the reduction of the mental to the physical in this way, Anomalous Monism, on one hand, allows the autonomy of the mental, while it keeps a physicalistic ontology on the other hand. However, Anomalous Monism faces a difficulty.

3-5-3 The Problem about the Causal Efficacy of the Mental

A serious problem is raised against Anomalous Monism. According to Anomalous Monism, an event *c* causes another event *e* if and only if the pair <*c*, *e*> is subsumed under a strict physical law. In case of mental causation <*m*, *e*> (where *m* is a mental event and *e* is a physical or mental event), *m* causes *e* in that m has a physical property *P*, and that *P* is connected with some physical property of *e* by a strict physical law. Therefore, it seems that the mental property *M* of *m* does not contribute to the causal relation at all in bringing about *e*. Suppose we accept the following principle:

> **The Principle of the Causal Efficacy of Properties (EP)**: For every event *c* and *e*, if *c* causes *e*, then there exists a property *F* of *c* and *c* causes *e in virtue of F*.

[16] Cf. Nagel (1961).

Then we must say that in the case of mental causation, the mental cause *m* caused the effect *e* in virtue of its physical property *P*, not its mental property *M*.[17]

Davidson himself replies to this objection as follows. As he sees it, causal relation is extensional relation. If an event *c* causes another event *e*, then this causal relation obtains no mater how these events are described: *c* causes *e simpliciter*. An event does not cause another event in virtue of something (physical or mental properties).[18] Therefore, he denies EP.

Davidson's reply, however, is not persuasive. First, there are cases that seem to support EP. Suppose a soprano singer shouted in a high note, "Shatter!", and thereby broke a glass window. Which property of her voice broke the window, the property of meaning 'Shatter!' or the property of having a high frequency? It seems to be the property of having a high frequency that is causally efficacious or active in bringing about the shattering of the window. The reason is this. Were the frequency of her voice not high enough, her voice would not break the window, which suggests that the property of having a high frequency is causally active. On the contrary, even if she sang a different word, it would nonetheless break the window if only its frequency is high enough. These considerations suggest that the property of having a high frequency has causal efficacy, while the property of meaning something does not.[19] We need, of course, to examine the notion of causal efficacy more precisely, but we can at least say as follows. In general, when an event which has various properties causes another event, it is one of the properties possessed by the event that is causally relevant; it is not true that all the properties are causally relevant, nor is it true that none of the properties is causally relevant. (We can call the property which is relevant in a causal relation 'the property which has causal efficacy'.) Denying EP, we could not predict or explain the occurrence of events based on the other events. This does not conform to the fact that we usually do these predictions and explanations. We can also point out the following: if the concept of causation cannot accommodate our ordinary practices such as predictions and explanations, it would be too vacuous, almost meaningless, as a concept of causation.

---

[17] Many philosophers point out this problem. See, for example, Stoutland (1980), Honderich (1982), Sosa (1984), and Kim (1984).

[18] See Davidson (1993).

[19] The example of a soprano singer is borrowed from Yablo (1992).

Second, we should notice that there is an asymmetry between mental descriptions and physical descriptions even in Davidson's own view. Why does he cite and accept the principle (P2), the Principle of the Nomological Character of Causation, which insists that causal relations need to be backed up by strict causal laws? Why is the existence of strict laws a necessary condition of the existence of causal relation? That is, as I see it, because Davidson himself concedes that laws and properties (descriptions) do a crucial role in causal relations. If mental properties (descriptions) do nothing important in causal relations, then Davidson must take the objection seriously.

The objection so far does not reproach Davidson for taking it that causal relata are events as particulars: the ontological item that serves as causal relata might be events as particulars, not properties or properties instances as some philosophers insists it. The objection insists that some *property* of a cause event must have causal efficacy in bringing about an effect event, and that Anomalous Monism cannot exclude the possibility that mental properties lose their causal efficacy.

3-5-4 Attempts to Save Anomalous Monism

As argued above, Anomalous Monism faces the problem that it cannot exclude the possibility that mental properties lack their causal efficacy. There have been several attempts to save Anomalous Monism from the objection. The notable attempts are as follows:

Supervenience: If we establish some relation between mental properties and physical properties that is not too strong to allow reduction of the former to the latter, we can give causal efficacy to mental properties.

Non-strict Laws: Although we don't have 'strict' mental laws, we have 'non-strict' mental laws and these 'non-strict' laws are sufficient for mental properties to have causal efficacy.

Counterfactual Dependence: Even though we don't have any mental law, we can at least find out a counterfactual dependence between mental properties and physical properties. This counterfactual dependence gives causal efficacy to mental properties.

Explanatory Practice: In our daily lives, we often explain our behaviours by mentioning our mental states. This explanatory practice should be the ground to give causal efficacy to mental properties.

In this section, I will only examine the first attempt, **Supervenience**. The other three, I will examine in the next chapter when I discuss the causal exclusion problem.

The main reason why Anomalous Monism cannot give causal efficacy to mental properties is because it requires a strict law to back up a singular causal relation and the strict law must be a physical law. However, there is another reason (thought it is related to the first reason): Anomalous Monism cannot give causal efficacy to mental properties because it does not specify any connection between mental properties and physical properties.

It would be plausible to suppose that the principle EP entails the following principle:

Counterfactual Test for Causal Efficacy (CT): For every event *c* and *e*, and every property *F*, if *c* caused *e* in virtue of *F* of *c*, then the following counterfactual conditional is true: ceteris paribus, if *c* hadn't had *F*, then *c* wouldn't have caused *e*.

Remember the example of a soprano singer. When we discussed the causal efficacy of the property of having a high frequency and the property of meaning something, we regarded the former as having causal efficacy because the voice wouldn't have broken the window if it had not had a high frequency, whereas even if the voice had not meant 'Shatter!', it would still have broken the window as long as the frequency was kept high enough. This consideration obviously relies on CT. Here, it is important to notice the following: CT insists that the truth of the counterfactual conditionals is a *necessary* (not sufficient) condition for causal relation. If a property does not satisfy CT, then the causal efficacy of the property is called in question. Even if a property satisfy CT, it is not sufficient to guarantee the causal efficacy to the property, as there is a case of a common cause producing two independent effects (I will discuss this problem later).

Given CT, it should be clearer why Anomalous Monism cannot give causal efficacy to mental properties. Anomalous Monism allows any physical events to have

any mental properties. Because of this generosity, mental properties cannot satisfy CT. It can be showed as follows:[20]

(1) Suppose that a mental event *c* caused another event *e*.
(2) By (P2), *c* and *e* must be subsumed under a strict law, and by [AM] it must be a physical law: a physical property *P* of *c* appears in the strict physical law.
(3) By (P3), a mental properties *M* of *c* cannot covary with a physical property *P* of *c*: it is always possible that *c* has *P* but lacks *M*.
(4) Therefore, a counterfactual conditional, 'ceteris paribus, if *c* hadn't had *M*, then *c* wouldn't have caused *e*', turns out to be false.

Anomalous Monism does not satisfy CT, therefore it cannot give the causal efficacy to mental properties.

This consideration gives us a clue to regain the causal efficacy of mental properties: we need to constrain the relation between mental properties and physical properties. Some philosophers, including Davidson himself, try to constrain the relation between mental properties and physical properties by means of *supervenience*. Davidson defines supervenience as follows:

> [A] Predicate *p* is supervenient on a set of predicates *S* if and only if *p* does not distinguish any entities that cannot be distinguished by *S*.[21]

This definition is equal to the Weak Supervenience, which is formulated by Kim as follows:

> [With *A* and *B* being families of properties,] *A weakly supervenes* on *B* if and only if necessarily for any *x* and *y* if *x* and *y* share all properties in *B* then *x* and *y* share all properties in *A* – that is, indiscernibility with respect to *B* entails indiscernibility with respect to *A*.[22]

---

[20] I am helped by Mino (1995) for the formulation of the following argument.
[21] Cf. Davidson (1993), p.4.
[22] Cf. Kim (1984), p.58.

The weak supervenience certifies that two objects sharing all properties in *B* will share all the properties in *A* just in case they are both in the same world. When two objects inhabit the different possible worlds with each other, the weak supervenience does not certify that the two objects sharing all properties in *B* will share all the properties in *A*. Because of this weakness, the weak supervenience cannot endow the causal efficacy to mental properties. The weak supervenience allows a possible world that is exactly the same as the actual world except that *c* lacks *M*. Therefore, a counterfactual conditional 'ceteris paribus, if *c* hadn't had *M*, then *c* wouldn't have caused *e*' is still false.

This shows that the weak supervenience is too weak to endow the causal efficacy to mental properties. Kim suggests the stronger version of supervenience which is formulated as follows:

> *A strongly supervenes* on *B* just in case, necessarily, for each *x* and each property *F* in *A*, if *x* has *F*, then there is a property *G* in *B* such that *x* has *G*, and *necessarily* if any *y* has *G*, it has *F*.[23]

However, as Kim sees it, this strong supervenience is so strong that it allows a kind of reduction of mental properties to physical properties.[24] Therefore, non-reductive physicalist cannot take this option. In sum, if supervenience is weak enough to keep non-reductiveness of mental properties, then mental properties cannot have causal efficacy; if, on the other hand, supervenience is strong enough to endow causal efficacy to mental properties, then mental properties are, in some sense, reduced to physical properties and they lose their independent reality.

3-6 The Principle of Physical Realization

The requirements are intended to form a *minimal* physicalism, so we cannot say that non-reductive physicalism is qualified as 'full-fledged' physicalism just because it satisfies these requirements. Some non-reductive physicalists might insist that the

---

[23] Cf. Kim (1984), p.65.

[24] See, for example, Kim (1989) p.283 and Kim (1998). We will see later what kind of reduction the strong supervenience allows.

second requirement, The Principle of Physical Determination, is too weak for the requirement for physicalism, as long as it is formulated by the concept of supervenience. They might suggest the following requirement for substitution:

> **The Principle of Physical Realization**: Mental properties must be realized by physical properties.

As it will become clear later, I will take this principle. However, it is important to notice that even if we take the Principle of Physical Realization, there could still be several options corresponding to which formulation of realization one takes. Jaegwon Kim, for example, formulates physical realization as follows[25]:

> A property *F* is physically realized by a property *G* just in case:
> (1) *F* is a second-order property over set **B** of base (or first-order) properties in that *F* is the property of having some property *P* in **B** such that *D(P)*, where *D* specifies a condition on members of **B**.
> (2) *G* is a member of **B**.

Although I accept the Principle of Physical Realization, I will not take Kim's formulation of realization. I will explain the reason why I deny it later when I discuss multiple realization and formulations of physical realization.

[25] Cf. Kim (1998), p.19-23.

4 The Causal Exclusion Problem

Non-reductive physicalism insists that mental causation is *real*. This should mean that mental properties are causally effective. Is it really possible? Jaegwon Kim offers an argument against non-reductive physicalism in quite a general form. It is called 'the causal exclusion argument'. The problem posed by the argument is called 'the causal exclusion problem' (or simply, 'exclusion problem'). I will, first, reformulate the argument in his *Mind in a Physical World* (p.38-47). Then, I will examine notable defences to the exclusion argument that have been offered so far. It will be argued that none of the defences are successful.

4-1 Kim's Attack on Non-reductive Physicalism: The Causal Exclusion Argument

Consider a case of mental causation where a mental event c at t_1 causes a physical event e at t_2 ($t_1 < t_2$). As c is a mental event and e is a physical event, c has a mental property M_1 at t_1 and e has a physical property P_2 at t_2 respectively. Then in this case, we can say that c's having M_1 causes e's having P_2. Let us suppose, for simplicity, that c's having M_1 is a sufficient cause of e's having P_2.

According to the Principle of Realization (or the Principle of Supervenience), which is required for a view to be physicalism, M_1 must be realized by (or supervene on) a physical property at time t_1 (let us call it P_1). That is to say, c's having P_1 determines c's having M_1, and c's having P_1 is sufficient for c's having M_1. Therefore, it turns out to be that c's having P_1 is sufficient for e's having P_2. Now, according to the Principle of Physical Closure, a physical event (e's having P_2, in this case) must have a sufficient physical cause, which is, in this case, nothing but c's having P_1. All the inference above, then, will have a very annoying consequence that what does the causal work in bringing about the physical effect, e's having P_2, is not the mental property M_1 but the physical property P_1: the mental property M_1 does not, in fact, have causal efficacy in the purported case of mental causation. Do we avoid this annoying

consequence for non-reductive physicalism?

There seem to be only two possible ways to avoid the consequence. First, we might be able to think that *c*'s having P_1 causes *c*'s having M_1, and then *c*'s having M_1 causes *e*'s having P_2. The second way would be such that both *c*'s having P_1 and *c*'s having M_1 are sufficient to bring about *e*'s having P_2. The first way regards the realization relation as a kind of causal relation. However, it is difficult to regard the realization relation as a causal relation. First, the realization of M_1 by P_1 should be considered to be simultaneous, whereas the cause is usually considered to precede the effect. There are two fundamental theories in contemporary physics: the relativity theory and the quantum theory. While quantum theory admits, in one sense, simultaneous causation, the relativity obviously denies it. It is an open question in current physics whether we can admit simultaneous causation or not. Therefore, we should not expect that we can assess the possibility of simultaneous causation just by a priori argument. Nor should we presuppose simultaneous causation in discussing the metaphysical problem of mental causation. Second, 'it is difficult… to imagine a causal chain, with intermediate links, between the subvenient and the supervenient properties. What intermediary events could causally connect a mental event with its subvenient physical base?'[26] Finally, to regard the realization of M_1 by P_1 as a causal relation might violate the Principle of Physical Closure. This is because *c*'s having P_1 must take a detour (i.e. causing *c*'s having M_1) in causing *e*'s having P_2.

The second way is to accept the overdetermination of the effect by both the mental cause and the physical cause. If we accept this, we must accept for *every* case of mental causation that the physical effect that is caused by the physical cause is also caused by the mental cause as well. A typical example of the overdetermination is the case of double assassins: two independent assassins shot the victim to death, two bullets hitting the heart of the victim simultaneously. While one bullet would be sufficient to kill the victim, he was unluckily shot double. This scenario, however, is not applied to mental causation. According to the Principle of the Physical Closure, for every physical effect, there must be a sufficient physical cause (if there is a cause at all). It follows from this that for *every* case of mental causation, the effect is always overdetermined by the physical cause and the mental cause. Overdetermination, if it happens, is considered

[26] Kim, *Mind in a Physical World*, p.44.

to happen accidentally. If non-reductive physicalist must require that overdetermination *always* happens in mental causation, it must explain this fact in some ways.

Non-reductive physicalism might, still, try to explain why an effect is always overdetermined in mental causation. It might be explained by linking the metal with the physical in some way. In fact, according to the Principle of Supervenience (or the Principle of Realization), *c*'s having P_1 determines *c*'s having M_1. A non-reductive physicalist might insist that because of this linkage, the overdetermination does not happen accidentally in the case of mental causation. But then, the next problem takes place. The physical cause, *c*'s having P_1, must be a sufficient cause of *e*'s having P_2, as it is required by the Principle of Physical Closure. Then what work is left for the mental cause in bringing about the physical effect? Isn't it *redundant*?

This is the reformulated outline of Kim's attack on non-reductive physicalism. The upshot of his attack is that in the case of mental causation, every mental property is pre-empted by the physical property which realizes the mental property in causing the effect, in the framework of non-reductive physicalism. And nothing is left for the mental properties to do in the mental causation. If so, the mental properties should be excluded by the physical properties that realize them in the case of mental causation. Kim called this problem the causal exclusion problem.

4-2 Some Defences from the Non-reductive Physicalist Camp

Many defences have been proposed from the non-reductive physicalist camp. In this section, I will overview some of the leading defences. It will be shown that none of the defences is successful.[27]

Defence 1: Counterfactual Dependence as a Ground for Causal Relation

The most popular defence tries to insist that the counterfactual dependence obtains between the mental properties of cause events and the properties of effect events, and

[27] The survey and the discussions in this section are based on the chapter 3 of Kim (1998) and Mino (2004).

the counterfactual dependence grounds the causal relation between them.[28] Suppose an event *c* has a property *F* and event *e* has a property *G*. This defence has it that we will have enough evidence that *c*'s having *F* caused *e*'s having *G* when the following two counterfactual conditionals are true[29]:

(1) If *c* had not occurred, then, other things being equal, *e* would not have occurred.
(2) Given that *c* did occur, the other things being equal, *e* would inevitably have occurred.

Consider, for example, a case of mental causation: I opened the window because I wanted to breeze fresh air. In this case, if I had not a desire to breeze fresh air, then, other things being equal, I would not have opened the window. Also, given that I had a desire to breeze fresh air, the other things being equal, I would have opened the window. Both (1) and (2) being satisfied, we are certified to say that the mental property (my desire) did cause the opening of the window and has causal efficacy.

However, there is a counter example to this defence. It is the case where one common cause brings about two different effects. Suppose that an event *c*, a common cause, brings about two different events, e_1 (with a property *F*) and e_2 (with a property *G*). If e_1 had not occurred, then, other things being equal, e_2 would not have occurred as *c* had not occurred either in this case (counterfactual conditional (1) obtains). And given that e_1 did occur, the other things being equal, e_2 would inevitably have occurred as *c* must have occurred in this case (counterfactual conditional (2) obtains). Therefore, a proponent of Defence 1 must say that e_1 caused e_2 and that *F* has causal efficacy in bringing about *G*, which is obviously absurd. What has causal efficacy in bringing about *G* is a property of *c*. This counter example shows that satisfying (1) and (2) is not sufficient to endow causal efficacy to a property.

A defender might try to revise the definition of causation to avoid the above counter example. Although I cannot take it in more detail here, it is very likely that counterfactual criterion is not appropriate for fully grasp the concept of causation. Anyway, we can at least say that the prospect of this approach is still open to

[28] See, for example, Baker (1993) and Horgan (1997).
[29] See Baker, 'Metaphysics and Mental Causation', p.93

discussion.

Defence 2: Non-Strict Mental Laws

The second defence insists that some causal laws include mental predicates and this guarantees the causal efficacy of the mental properties indicated by the mental predicates appearing in the causal laws. The point is, causal laws do not have to be strict laws; a non-strict law with ceteris paribus condition would be fine for it to guarantee causal efficacy. Jerry Fodor, for example argues like this. If there is a causal law such that ceteris paribus, an event with *F* causes another event with *G*, then a property *F* determines which event the event with F causes. This is to say that *F* is the property in virtue of which an event with *F* causes another event. Therefore, we must say that *F* is the property with causal efficacy. In general, a property *F* has causal efficacy if and only if there is a law such that ceteris paribus, an event with *F* causes another event with *G*. The ceteris paribus condition in the formulation implies that the causal law might be a non-strict one, but this does not spoil the causal efficacy of *F*. That is because the property *F* can determine which event the event with *F* can cause, *given* that ceteris paribus condition obtains. Now, we do have causal laws (perhaps with ceteris paribus conditions) including mental predicates, therefore the mental properties corresponding to those mental predicates are causally efficacious and real properties.[30]

This defence, however, has the following problem.[31] Suppose that a pair of events <*c*, *e*> is subsumed under a mental law such that ceteris paribus an event with *M* (a mental predicate or property) causes another event with *P* (a physical predicate or property), and suppose also that ceteris paribus condition obtains. Is this pair <*c*, *e*> subsumed under some physical law as well? If the answer is no, then the Principle of Physical Closure will be violated. Because the fact that there is no physical law subsuming <c, e> implies that *e*'s having *P* does not have a physical cause. On the other hand, if the answer is yes, then we must say that there exists a physical law subsuming <*c*, *e*> such that ceteris paribus an event with *P*' causes another event with *P*. This means that the property *P'* determines the causal relation of <c, e>. However, this

[30] See Jerry Fodor, *A Theory of Content and Other Essays*, p.152.
[31] Cf. LePore and Loewer (1989).

defence also insists that the property *F* determines the causal relations of <*c,e*>. Here the overdetermination kicks in again. If we don't want to violate the physical closure in this state, we must render the mental property redundant. A dilemma of this sort occurs again and again in the following.

Defence 3: Generalization of Mental Causation

Defence 3 insists that if mental causation does not exist, then causations in special sciences, such as chemical causation and biological causation, do not exists either, which is quite ridiculous. The argument goes like this. Suppose that mental properties don't have causal efficacy and what really do the causal works in purported 'mental causation' are neuro-physiological properties which realize mental properties (or which mental properties supervene on). However, those neuro-physiological properties are also realized by lower-level properties such as physical or chemical properties. And those physical or chemical properties are realized by lower-level properties such as micro physical properties. This realization process will not end until it comes to the ultimate physical properties (if there are such properties at all) that will be specified by future fundamental physics. If the causal efficacy of the mental properties is pre-empted by the lower properties that realize them, then all the properties other than the ultimate physical properties, the properties such as macro physical properties, biological properties, or chemical properties, would be pre-empted by the ultimate physical properties as well, and would lose their own causal efficacy. We would, then, be unable to literally state that a hitting of a glass caused the glass to shatter or that a gust of strong wind caused a tree to fall down: these statements would not be literally true. This is absurd. Therefore, by *reductio ad absurdum*, we must say that metal properties have causal efficacy and reality. Lynne Rudder Baker, a notable philosopher trying to defend non-reductive physicalism in this line, insists as follows:

> Moreover, I want to show that the metaphysical assumptions with which we began inevitably lead to scepticism not only about the efficacy of contentful thought, but about macro-causation generally. But if we lack warrant for claiming that macro-properties are generally causally relevant, and if we take explanations to

> mention causes, then most, if not all, of the putative explanations that are routinely offered and accepted in science and everyday life are not explanatory at all. (Baker, 'Metaphysics and Metal Causation', p.77)

Many philosophers try to defend the causal efficacy of mental properties in this line.[32]

Defence 3 might, however, mistakenly state that when a property is realized by another property, the realized property and the realizing property belong to the different levels in macro-micro hierarchy. To take a biological property, being a gene, as an example, the defence might have it that being gene is realized by a lower level chemical property such as being a base sequence, and that the latter is a micro property, whereas the former is a macro property. However, as Kim points out rightly, this is not a right way to see the situation[33]. The property that realizes a macro property is not a micro property as Defence 3 insists. The property that realizes a biological property, being a gene, is the base sequence of the *aggregate* of bases. Kim calls it (in this case, the property that realizes a biological property) a micro-based property and formulates it using David Armstrong's notion of 'structural property' as follows:

> *P* is a *micro-based property* just in case *P* is the property of being completely decomposable into nonoverlapping proper parts, a_1, a_2, … , a_n, such that $P_1(a_1)$, $P_2(a_2)$, … , $P_n(a_n)$, and $R(a_1, \ldots, a_n)$. (Kim, *Mind in a Physical World*, p.84)

Using Kim's term, we should say that a micro-based property is not a micro property but a macro property. The exclusion problem concerns two (or more) properties in the same level. Therefore, Defence 3's worries that the causal efficacy of a macro property (or a property belonging to 'higher level') is pre-empted by that of a micro property (or a property belonging to 'lower level') and that the causal efficacy drains away, miss the point of the exclusion argument.

Of course, this is not to insist that there is no problem about the causal efficacy

---

[32] See, for example, the following studies: L. R. Baker, 'Metaphysics and Mental Causation'; Tyler Burge, 'Mind-Body Causation and Explanatory Practice'; Robert van Gulilck, 'Who's in Charge Here? And Who's Doing All the Work?', all in Heil and Mele, eds., *Mental Causation*, 1993.

[33] See Kim (1998), p.80-87.

of the properties in special sciences. Quite the opposite. If all the macro properties and all the properties in special sciences supervene on (or realized by) the micro-based properties composed of the properties in the ultimate physics (and it is plausible enough), then the causal efficacy of those macro properties and the properties in special sciences would be pre-empted by the micro-based properties. And all the macro properties and the properties in special sciences are to be either not causally efficacious or redundant. But, again, we must notice the following: it is not that the exclusion argument has an absurd consequence that macro properties and properties in special sciences lose their causal efficacy, but rather that the exclusion argument asks what kind of ontology we could take in the physicalistic framework, keeping our intuition that there exists mental causation. To solve the exclusion problem for mental properties is also to solve the problem for other macro properties and properties in special sciences.[34]

Defence 4: Causing (or Explaining) Different Effects

One might insist that a mental explanation and a physical explanation does not explain one and the same effect but, in fact, two different effects. We can explain an agent's behaviour of opening the refrigerator by referring to the agent's mental property, the desire to drink beer and the belief that there is a bottle of beer in the refrigerator. We can also explain the same behaviour by referring to the agent's physical property (the neuro-physiological property of the agent's brain). However, as this defence insists it, these explanations do not, in fact, explain the same effect. While the mental explanation explains the *action* of the agent, the physical explanation explains the bodily (physical) *behaviour* of the agent.[35]

If we can regard two explanations as explaining one and the same event or the different events depends on the ontology of events. If we take Kim's view on events, we can regard two explanations explain two different events. If we take Davidson's view on events, we would regard the explanandum of two explanations are two different properties of one and the same event. Anyway, it seems that we do not have to worry

[34] Cf. Kim (1998), pp.77-87.

[35] Amie Thomasson offers a defence of nonreductive physicalism in this line. See her 'A Nonreductivist Solution to Mental Causation', in *Philosophical Studies* 89, 1998, pp.181-95.

the overdetermination any more.

It is true that according to this defence, we do not have to worry about the overdetermination of the effect. However, it is not possible to secure the causal efficacy and the reality of mental properties in this line. First, according to this view, we must admit that every mental causation would be mental-mental causation. But isn't it counter-intuitive that our mental states or properties cannot cause our behaviours as physical events? We usually believe that our mental states can sometimes cause our behaviours. Event though the mental event (the action) which is caused by the mental property turns out to be a physical event (an event described in both mental and physical predicates) in Davidson's event ontology, it would not very happy for non-reductive physicalism to give up the intuition that mental properties can cause physical events as physical.

Second, and more importantly, it seems that the mental explanation that explains actions by referring to mental properties is not itself a 'causal' explanation any more. As I have already showed when discussing Anomalous Monism in 3.5, the mental properties (or descriptions) of events are in the relation of rationality: they are connected by normative principle. Of course, as Davidson puts it, for a mental event to be *the* cause of the action, it is not sufficient for the mental event is a reason of the action: the mental event must cause the action. However, the mental event causes the action in virtue of its physical description appears in the strict causal law that connect the event with the action (the effect). Therefore, mental properties does not have any causal efficacy by its own. With this consideration, I insist that the explanation by referring to an agent's mental properties is not a causal explanation by itself. Therefore, I also insist that non-reductive physicalist cannot secure the causal efficacy and the reality of mental properties in this line.

Defence 5: Taking our Explanatory Practice Seriously

Some defenders of non-reductive physicalism argue, against Kim's objection, that we can save the reality of mental causation by taking our *explanatory practice* seriously. It goes like this. The causal exclusion argument takes an ontological principle, the Principle of Physical Closure (in some form or other), seriously. It argues from the

ontological principle to the conclusion that we must deny our common activity of explaining our behaviour by referring to our mental properties or mental states. But this is, so the defender insists, the other way around. We normally refer to mental properties in the causal explanations of behaviours. This type of explanation is widely conducted and widely accepted as a proper explanation. We must not, therefore, deny it so easily. The right way to take is to take our explanatory practice seriously, thereby securing the causal efficacy of mental properties. As Tyler Burge puts it, 'materialist metaphysics is not the most plausible starting-point for reasoning about mind-body causation. Explanatory practice is.'[36] If we refer to a mental property in a causal explanation of behaviours and the causal explanation is plausibly considered to be a true (or a proper) explanation, then the mental property should be regarded as a real property.[37]

However, if this defence regards the causal exclusion argument as a kind of sceptical argument that casts doubt on the existence of mental causation, then the defence misses the point. That is because the point at issue is not an *epistemological* problem, the problem concerning whether our assertion that there exist mental causations is well grounded, but rather an *ontological* problem concerning the existence of mental causation. We begin by noticing the fact that our belief in mental causations is firm, and then go on to argue how the mental causation is *ontologically possible* in the framework of non-reductive physicalism. The exclusion argument insists that non-reductive physicalism is not a proper ontological theory, based on the consideration that the existence of mental causation is not compatible with the ontological framework of non-reductive physicalism.

The point of the defence might be that there could exist two causal chains (or processes), both of which are equally qualified as real properties and are compatible with each other, leading to a physical event. Tyler Burge insists:

> I think that we have reason … to think that mentalistic and physicalistic accounts of causal processes will not interfere with one another.

---

[36] Tyler Burge, 'Mind-Body Causation and Explanation', in Heil and Mele, eds., *Mental Causation*, 1993, p.118.

[37] See Tyler Burge, 'Mind-Body Causation and Explanation', and L. R. Baker, 'Metaphysics and Mental Causation', both in Heil and Mele, eds., *Mental Causation*, 1993.

> … we know that the two causal explanations are explaining the same physical effect as the outcome of two very different patterns of events. The explanations of these patterns answer two very different types of inquiry. (Burge, 'Mind-Body Causation and Explanation' p.115-6)

So, the Defence 5 might insist that a mental explanation (explanation referring to an agent's mental state) and a physical explanation (explanation referring to an agent's bodily state) of the agent's behaviour are compatible with each other and that both the agent's mental state and the agent's bodily state could equally be a cause of the agent's behaviour.

A question, however, immediately arises for this insistence: what is the relation between the two causes, the mental state and the bodily state of the agent? If each cause, independent with each other, is sufficient for the behaviour, then the problem of overdetermination arises again: we must either assume the massive coincidence of overdetermination or admit that one of the causes (i.e. the mental cause, because of the Principle of Physical Closure) is redundant. Or is one of the causes is a part of the other? But if one takes 'non-reductive' physicalism, one would insist that mental properties are different from physical properties, and therefore would not like to take this option. Is it, then, possible to think that two causes are different parts of a whole cause which is sufficient for the behaviour? This is the case when both striking a match and the existence of enough oxygen are different parts of a whole sufficient cause for the lightening of the match. However, we cannot apply this case to mental causation if we try to stick to non-reductive physicalism. If mental causation is a case of this kind, then a physical cause (the bodily state of the agent) turn out to be insufficient for bringing about the behaviour: it needs a help by the mental state to cause the behaviour. This is a violation of the Principle of Physical Closure.

These are, of course, ontological questions, and the defender might insist that we cannot presuppose the plausibility of ontological considerations. However, the point is this. Two explanations, each of which is different from the other in that they offer us different information about the object (or the world), could nonetheless be compatible with each other. But if they are causal explanations and their explanatory powers come from their describing some aspects of the world, then we cannot avoid the ontological

questions concerning the conflict between two causes. We must remember that the problem at issue is the reality of mental properties and mental causation: this is an ontological matter.[38]

4-3 Can We Weaken the Principle of Physical Closure?

It would be fairly obvious now that the Principle of Physical Causal Closure plays a crucial role in the exclusion argument. As long as it is required as a necessary condition for the physicalism, any property which is distinguished (in some sense) from physical properties is always exposed to the danger that it gets deprived of its causal efficacy and therefore its reality as well.

Can't we, then, secure the causal efficacy of mental properties by *weakening* the Principle of Physical Causal Closure some way or other? Eric Marcus distinguishes two versions of the principle; *Completeness* and *Closure*[39]:

> According to Completeness, all physical events have complete physical causal histories. …
> … According to Closure, physical events cannot interact causally with non-physical events, or with physical events in virtue of their non-physical properties. (2005, p.28-9)

Completeness, a weaker version, has it just that when we trace back a sufficient cause of a physical effect, we *need not* get outside the physical domain: the physical domain is complete in that the network of the physical is self-contained without any help of the non-physical. On the other hand, Closure, a stronger version, has it that the non-physical cannot get into the causal relations with the physical: it prohibits the non-physical from getting into the physical causal network. Marcus insists that we are not forced to accept Closure: it is too strong and it needs some justification that has not been offered yet. On the other hand, continues he, Completeness is plausible enough.

---

[38] For more discussion, see Kim (1998), pp.60-67.
[39] See Eric Marcus (2005), p.27-50.

We should now notice that Completeness itself does not exclude the non-physical. It just says we need not get outside the physical domain. It does not prohibit the causal histories of the physical effects getting outside the physical domain.[40]

Does Marcus succeed in solving the exclusion problem by weakening the principle? In earlier part of this thesis, section 2-4, I formulated the Principle of Physical Causal Closure as follows:

> **The Principle of Physical Causal Closure**: For all physical objects, if it has a cause at *t*, it has a sufficient physical cause at *t*.

It is quite obvious that our Principle of Physical Causal Closure is virtually identical with Marcus's Completeness, which Marcus himself is willing to accept. However, in the preceding part of this chapter, we saw a consequence that mental causes are excluded by the physical cause, which Marcus does not accept. Why does this happen? In what point do Marcus and I depart?

The point of departure is that he accepts redundancy and I do not. It is true that the Principle of Physical Causal Closure does not in itself prohibit mental properties to get into physical causation. But if we don't have any clue to understand the correlation between mental properties and physical properties, then we must just accept the mysterious coincidence. On the other hand, if we think that mental properties are in some sense dependent on physical properties, then it seems that mental properties are redundant. Physical properties do all the work. Why we need mental properties in addition to physical properties? If the reason is just to save our intuition about the reality of the mental, it seems quite *ad hoc* to me to just introduce mental properties by this reason.

[40] Ibid.

5 Kim's Reductionism

In this chapter I will review and examine Jaegwon Kim's solution to the exclusion problem. The upshot of Kim's attack on non-reductive physicalism is that non-reductive physicalism is quite an unstable ontological position. That is to say, the non-reductionist intuition and the physicalist intuition are not compatible. If mental properties are to be reduced to physical properties, then one would naturally take 'reductive' physicalism ('outright' physicalism); if mental properties are not to be reduced to physical properties, then one must either give up physicalistic intuition and take a dualistic view or otherwise take eliminative or epiphenomenal physicalism.

5 - 1 Kim's Solution to the Exclusion Argument

Kim's solution is to take a reductionist way. How, then, mental properties are reduced to physical properties? We have already seen that the case of the multiple realizability of mental properties placed the reduction of mental properties in serious trouble. Kim tries to accomplish the reduction by revising 'the concept of reduction' itself.

As Kim sees it, the way to a new theory of reduction is suggested by functionalism.[41] According to functionalism, mental properties are functional properties which are realized by physical properties. In other words, a mental property is described as a causal role which a physical property occupies (or plays). More precisely put, a mental property is a second-order property, a property of having a first-order physical property that satisfies some causal condition. For example, a mental property of having a pain is a property of having a first-order physical property that satisfies the causal condition such that typically it is caused by the tissue damage and causes the wincing behaviour.

What is a second-order property of having a first-order property in general? As Kim sees it, a second-order property should be nothing but *a property of having a first-order physical property*. For, when we say that a particular object has a

[41] For Kim's 'functional reduction', see Kim (1998), p.19-27, 97-101.

second-order property, the truth-maker of this statement should be nothing but the first-order physical property that realizes the second-order property that makes the statement true. An object having a second-order property is nothing more or less than its having the first-order property that realizes the second-order property. Therefore, if a mental property is a second-order property (functional property), then it would be identified with the first-order physical property that realizes the mental property. In this way, we can say that if a mental property can be interpreted as a functional (functionalizable) property, then we would be able to reduce it to the physical property.

What about the multiple realizability? As we have already seen, a mental property, having pain, might be realized by P (e.g. having C-fibre activation) in human beings and by some quite different physical property Q in octopus. Then, we cannot identify a mental property of having pain with P. That is because, if we identify pain with P, then in quite the same way we can identify pain with Q as well, and by transitivity of identity, we would have to say that P is identical with Q; this is, of course, absurd.

Can we, then, identify pain with a *disjunctive* property of $P_1 \vee P_2 \vee \ldots \vee P_n$, where $P_1, P_2, \ldots, P_n$ are physical properties that realize a mental property pain in each of different systems or organisms. But this is quite implausible. Consider, for example, the property of having a primary colour.[42] This second-order property is realized by any of the first-order properties, having red, having blue, and having green. The proposal, however, would come down to insist that there is a disjunctive first-order property of having red or having blue or having green, in addition to the first-order properties such as having red, having blue, and having green, and this disjunctive first-order property is identical with the second-order property of having primary colour. But it is not plausible that there is such kind of disjunctive first-order colour. Therefore, in the same reason, we cannot identify a mental property of having pain with the disjunctive first-order physical property.

Facing the multiple realizability, in what way should a reductionist go? We have already confirmed that for every mental property, we cannot find a single physical property that is identified with it. As Kim insists it, we should give up the idea that when we use a mental predicate, there is one mental property corresponding to the

[42] This example is from Kim (1998) p.20.

predicate. Although there might be a single *concept* of pain, there is not a single *property* of having pain. These mental concepts or predicates refer to the different physical properties according to the different contexts. If a mental concept pain is applied to human beings, it refers to the physical property of having C-fibre activation. If applied to octopus, it refers to the physical property Q which is quite different from having C-fibre activation. And so on.[43]

Considering the multiple realizability of mental properties in this way, Kim offers a solution to the exclusion problem. It is quite simple: there is not a mental property that is multiply realized by various physical properties and excluded by them. There are just mental concepts or predicates. A mental concept or predicate refers to different physical properties in each case of the realization. For example, a mental predicate (or concept) 'having pain' refers to having C-fibre activation when it is applied to human beings, to a different physical property Q when it is applied to octopuses, and so on. Therefore, in each case of realization, what causes the physical effect is the first-order physical property that realizes the mental property.[44]

There could be some objections to Kim's reductive solution. Let us see some of them.

5-2 Doesn't It Deny the Autonomy of Psychology?

Jerry Fodor criticizes Kim for his denying the autonomy of psychology. He sums up Kim's view as follows:

> … Kim thinks philosophers haven't gotten it right about why MR states are ipso facto unsuitable for reduction. Once they do, Kim says, they'll see that the moral of 1&2 isn't, after all, that psychology is autonomous. Rather, it's that quotidian psychological states aren't reducible because they aren't projectible. Unprojectible states are, by definition, not the subjects of a possible science; they aren't bona fide kinds and they can't appear in bona fide nomological explanations. A fortiori, terms that express psychological states are not available for incorporation in "bridge laws"

[43] See Kim (1998) pp.103-112.
[44] Ibid.

or in (metaphysically necessary) property identities. This is all, of course, contrary to what a lot of philosophers, to say nothing of a lot of psychologists, have hitherto supposed. (Fodor, 1997, p.150)

Understanding Kim's view as above, Fodor criticizes it as follows:

> In effect, Kim wants to make it true *by fiat* that the only projectible kinds are physically homogeneous ones. …
>
> But, for better or worse, you don't get to decide this sort of thing by fiat ... Only God gets to decide whether there is anything, and, likewise, only God gets to decide whether there are laws about pains; or whether, if there are, the pains that the laws are about are MR. Kim's picture seems to be of the philosopher impartially weighing the rival claims of empirical generality and ontological transparency, and serenely deciding in favor of the latter. *But that picture won't do*. (ibid. p.161)

Fodor's criticism here has some connection with but is different from the defence 3 (the generalization of mental causation) discussed in the previous chapter. As Fodor sees it, Psychology, as well as the other special sciences, is autonomous from micro-physics in that mental properties appearing in several theories of psychology are *projectible*. The fact that mental properties appearing in the theories of psychology are projectible is , as Fodor insists, undeniable, because there are, in fact, many confirmed empirical generalizations in the theories of psychology. According to Kim's view, however, there do not exist such properties as multiply realizable mental properties. According to Kim's view, only mental predicates or concepts exist, and they refer to various physical properties in each distinct context where they appear. As Fodor puts it, Kim sticks to 'ontological transparency' of physicalism, and thereby denies well confirmed empirical facts – the projectibility of mental properties and the autonomy of psychology.

Kim, however, responds to this objection and ask: is the projectibility of mental properties so strongly confirmed? As Kim sees it, there are some empirical facts that seem to cast a doubt on the projectibility of mental properties. Using Kim's example, consider the following possible law: 'Sharp pains administered at random intervals

cause anxiety reactions'[45]. Suppose this generalization has been well confirmed for humans. Should we expect on that basis that it will hold also for other organisms that can have pains and whose psychology is implemented by a vastly different physical mechanism?[46] We would not expect it. That is because we can plausibly conceive an organism whose pain reactions we can observe, but whose anxiety reactions we cannot. We would not expect anxiety reactions of such organisms as octopuses, worms, and so on. We do *not* expect that the generalization about pains that has been well confirmed for humans will hold for some of the other organisms. In view of this example, Kim's denial of the projectibility of mental properties seems to have enough grounds, whereas Fodor needs more grounds for his view.

5-3 Doesn't It Come Too Close to Eliminativism?

There might be an objection that says Kim's view comes quite close to eliminativism about mental properties and it does not save the causal efficacy and the explanatory relevance of mental properties. Terence Horgan stresses this worry about Kim's view in a footnote of his paper as follows:

> But this turn toward eliminativism about mental properties appears to throw out the baby of mental realism along with the bathwater of psychophysical property-dualism. If there are no mental properties at all, then it is harder than ever to see how token mental events could be causally efficacious qua mental. Put another way, it is harder than ever to see how mentalistic causal/explanatory 'because'-statements, such as "She winced because she was thinking of Pat Buchanan," could ever be true. We do not save the causal efficacy and the explanatory relevance of mental properties by denying their existence. (Horgan, 1997, p.182)

This is certainly a good point and I largely agree with this objection. But before pursuing an alternative view from the next chapter, I would like to point out a couple of points.

---

[45] This example is from Kim (1992), p.324.
[46] ibid.

First, we might be able to distinguish Kim's reductionism from robust eliminativism such as one advocated by Churchland or others. It is true that in Kim's view, the mental properties as a second-order (functional) property do not exist: what exist are second-order predicates or concepts. However, those predicates or concepts *do* refer to some first-order physical properties, even though the referred physical properties are different in each case of realization. Those real physical properties which are referred to by mental concepts in each case are real, of course. And what is more, we might be able to say that those real physical properties are, at least in one sense, mental properties as well, because those real properties are truly referred to by mental predicates of concepts in each case of realization. Eliminativism, on the other hand, should be understood as the following view. For example, Churchland, a robust eliminativist, insists that folk psychology should be eliminated and the mental properties appearing in folk psychological theories should also be eliminated: he insists that the mental properties referred to in folk psychological theories are not real. The reason why he insists on eliminativism about folk psychology is as follows: first, there are many mental phenomena which folk psychology cannot explain at all, such as the nature of mental disorder, the nature and function of sleep, the ability to throw a snowball at moving car, etc.; second, folk psychology cannot be reduced to fundamental physical theories[47]. Therefore, Kim's view could be distinguished from eliminativism about mental properties insisted by, for example, Churchland. I think Kim holds a principle such as follows: whether one commits oneself to the reality of mental properties depends on whether one admits that mental predicates or concepts refer to some *real* properties even though they might be different in different cases of realization. If this principle is admissible, then we might be able to think that Kim is a realist about mental properties.[48]

However, we can at least say this: in Kim's view, mental properties do not have causal efficacy or reality *distinguished from that of physical properties*. Mental properties might be causally efficacious and real, but they are not causally efficacious or real *as mental properties* different from physical ones. I will go back to this problem

[47] See, for example, Churchland, P. "Eliminative Materialism and the Propositional Attitudes".

[48] See, for example, 'Postscripts on mental causation' in Kim (1993) for the related arguments.

later, when I discuss John Heil's view in Chapter 6 and David Robb's view in Chapter 8.

6 The Need for the Ontology of Properties (1): Towards a Causal Trope Theory of Properties

In this chapter and the next, I will depart from the problem of mental causation for a while, and focus on the ontology of properties. Overall, I will try to defend *a Causal Trope Theory* of properties, which is a disposition essentialist view combined with a 'sparse' trope theory. Its detail will be explained and clarified throughout this chapter and the next. Before beginning this chapter, however, I will briefly explain why I deal with the ontology of properties.

As we have just seen in the previous chapters, very few philosophers of mind have focused on the ontology of properties. It is obvious in the case of Donald Davidson. Davidson's Anomalous Monism insists that one and the same event can be *described* both by physical vocabulary and by mental vocabulary. It does not explicitly state that there are properties, as ontological entities, corresponding to all or some of the descriptions (physical or mental). Suppose an event can be described both as 'Alice's desiring for drinking orange juice' and as 'Alice's desiring for rehydrating herself'. Do these descriptions correspond to different properties? Or is it the case that an event which has a mental property of 'desiring for drinking orange juice' has *another* mental property of 'desiring for rehydrating oneself'? This kind of indifference of properties is observed in the works of some non-reductive physicalists as well. As was shown in the preceding chapters and as will be shown in the later part of this thesis, much confusion in the discussions on mental causation (and other problems in the philosophy of mind) arise from the confusion in the ontology of properties.

I will begin this chapter by focusing on John Heil's ontology of properties and substances, which is expounded in his book, *From an Ontological Point of View* (hereafter, *FOPV*)[49]. This will set the ground for me to expound my own view on properties, a Causal Trope Theory.

[49] Heil (2003).

6 - 1 Heil on the Levels Conception of Reality

John Heil's general message in *FOPV*, is as follows: many philosophers have attached too much importance to linguistic analysis rather than ontological considerations, and now we have to take ontology seriously; ontological considerations will help us resolve (or dissolve) many philosophical problems that have been besetting us. I am very sympathetic with this general message. What, then, is his view more specifically? What exactly does Heil mean by ontological considerations? What philosophical problems does he insist can be resolved by his ontological studies and how is it possible?

His main ontological claims which are argued for throughout *FOPV* can be listed as follows:

(1) The levels conception of reality is denied.
(2) There are no higher-level entities like human beings or horses, nor is it possible that talks about human beings or horses be linked analytically to truth-makers for such talks. Nevertheless, predicates like 'is a human being' or 'is a horse' are literally and truly applicable to particular objects.
(3) Properties are modes (tropes).
(4) All properties are both dispositional and qualitative at the same time.
(5) All properties are, strictly speaking, intrinsic properties.
(6) Heil lists *substances* in his ontological catalogue; therefore, any form of bundle theory is denied.

I will examine the first two claims in this section and the third claim in the next section. The remaining claims will be discussed in section 6-3, after I expound my own view on properties.

Let us start with (1), the denial of the levels of reality. The chapters from 2 to 7 of *FOPV* are devoted to deny 'the levels conception of reality'. When we think that our languages represent or *mirror* reality ('Picture Theory' in Heil's terminology), we tend to think the following principle (φ) holds:

(φ) When a predicate applies truly to an object, it does so in virtue of designating a

property possessed by that object and by every object to which the predicate truly applies (or would apply).[50]

Principle (φ) can lead to the idea that the world is composed of multiple layers. Take, for example, a predicate, 'is an egg beater'. Many types of physical devices can be told as 'an egg beater'. We can apply the predicate, 'is an egg beater', truly to a simple stick made of wires. We can also apply the same predicate truly to a more complicated electrical machine which is usually used in large factories. If the predicate, 'is an egg beater' is applied truly to all these objects in virtue of each objects having the same property corresponding to the predicate, then each object has both its own specific property and the property of *being an egg beater*. A simple stick has its own specific physical properties: it has a property of *being made of wires*, a property of *being 15cm*, and so on. The stick, however, has the property of *being an egg beater* as well if Principle (φ) holds. So does a more complicated electric egg beating machine. It is usually said that a functional property can be multiply realized by various properties. If principle (φ) is true, then we are easily led to the idea that the world consists of many layers – the ground layer of microphysical substances and properties, the upper layer of macro substances and properties, and so on. This is 'the levels conception of reality', the idea that Heil denies. He denies principle (φ), and thereby denies the existence of the upper layers. The alleged 'functional properties', he says, are not really properties but just *predicates*. It is a mistake to posit genuine properties and objects to all the predicates which truly apply to the world.

Here, we should notice that the levels conception of reality could mean two things: the levels of properties and the levels of objects. In case of predicates such as 'is an egg beater', Heil's claim amounts to the denial of the levels of properties. In case of predicates such as 'is a human being', it amounts to the denial of the levels of objects, substances, or kinds.

As argued in Chapter 3, multiple realizability has often been considered to support non-reductive physicalism in the history of the philosophy of mind. Non-reductive physicalism, as Kim persuasively argued, has a serious problem, the causal exclusion problem (see Chapter 4). If, however, we follow Heil and deny the

[50] *FOPV*, p.26.

levels conception of reality, the causal exclusion problem just *vanishes away*, for there is no multi-causations here! This perplexingly simple resolution of the causal exclusion problem has some similarity with Kim's solution[51] explained in the previous chapter.

We are inclined to ask the following question: is Heil's solution not a 'reductionistic' resolution or an 'eliminativist' resolution of the exclusion problem? This concerns the claim (2) of Heil's view. Is not the denial of the levels conception of reality a reductionistic view or eliminativistic view on higher-order properties (objects)? The answer depends on the meaning of the reduction and elimination. In Chapter 6, 'Philosophical Analysis', of *FOPV*, Heil argues as follows. Thinking about the nature of Fs, ordinary philosophers start with the question, 'Is there any analytical route to Gs?' There are two horns according to the answers to this question. If the answer is 'yes', then Fs are reducible to Gs (i.e. Fs are Gs). If the answer is 'no', then Fs are not Gs, where we have two choices: we might just eliminate Fs or we might accept that Fs exist *over and above* Gs. Such are the ordinary lines most philosophers tend to take. Heil, however, casts doubt on these lines of thought. According to Heil, even if there is no analytical route from Fs to Gs, it is possible to state *both* that Fs are not eliminated *and* also that Fs do not exist over and above Gs.

> There is little or no prospect of a systematic mapping between talk of statues and talk of collections of particles. But it need not follow either that, in addition to the particles, the universe contains statues, or that there are no statues. (*FOPV*, p.53)

Here we should distinguish ontological reduction and epistemological (or conceptual) reduction. Epistemological reduction entails ontological reduction, but not vice versa. When there is no analytical route about a mental entity, this mental entity is not conceptually reducible, but the mental entity still could be ontologically reducible to the fundamental entities. Heil's argument has the following scheme. Even if Fs are not analyzed to Gs, Fs are in some sense (i.e. in an ontological sense) equal to Gs, therefore we don't fall into eliminativism. As we have seen, Kim can be interpreted as taking a retentive reductionist view about the mental, with functional analysis model for

[51] As I will show later, Robb's solution also has a similar consequence. See Chapter 8.

reduction, which means that he keeps *conceptual* reduction in mind.[52] Heil, on the other hand, does not think that mental properties are conceptually reduced to fundamental properties. But, as Heil sees it, this conceptual irreducibility does not entail that mental properties do not exist (are eliminated) or they exist over and above fundamental (probably physical) properties. As Heil sees it, the conceptual irreducibility of mental properties is perfectly compatible with that they exist *as physical properties* (in other word, they are *ontologically reduced* to physical properties). There is obviously a similarity between their views.

There might be a slight difference between them, though. Kim thinks that phenomenological experiences (qualia) cannot be functionally analyzable and, therefore, they are irreducible to physical properties. In Kim's view, therefore, the irreducible qualia might raise the causal exclusion problem. In Heil's view, on the other hand, there is a prospect for ontologically (not conceptually) reducing qualia and, therefore, a prospect for avoiding the causal exclusion problem.

Anyway, setting aside the details, I agree with Heil in that we should not uncritically accept the levels conception of reality. Also, I have to say that I have a worry about Heil's view with respect to the causal exclusion problem. The worry is quite similar to what I mentioned about Kim's solution. In Heil's view, mental properties are ontologically reduced to physical properties although they are not analyzed by them. Mental properties are, as he sees it, not causally efficacious *as mental* properties but as physical properties. But do we really have to give up the causal efficacy of mental properties *as mental properties*? If we have another way to save the causal efficacy of mental properties as mental properties, we should, as I insist it, take that way. I will go back to this problem in Chapter 8.

6 - 2 Heil on Modes (Tropes)

Now let us go to the third claim of Heil's view: properties are modes (tropes). Heil is a robust particularist: he denies universals, and claims that properties are modes. Modes are 'particularized ways objects are'.[53]

---

[52] See Kim (1998) and Kim (2004).

[53] *FOPV*, p. 138. Heil prefers 'modes' to 'tropes' because, as he sees it, 'trope theorists

Heil insists that the decision between the two theories, a universalist theory and a trope theory, should be made on the basis of the benefits and costs of those theories.[54] As he sees it, the main benefit of the universalist theory is that 'universals promise a significant explanatory pay-off'[55]. First, if we have universals, we have a solution to the one-over-many problem. Second, if we have universals, we can easily explain the similarity relation among objects in terms of strict identity. Third, strict identity enables us to answer Hume's sceptical question: why should we expect similar things to behave similarly? That is to say, universals warrant inductive inferences, expressing our expectation that similar objects will behave similarly. The main cost of the universalist theory, on the other hand, is its counter-intuitiveness: a universal can be wholly present at many places at once. One more cost of the universalist theory might be that it is *less parsimonious*, as it posits an extra ontological item, universals, in addition to particular objects. As to a trope theory, the benefits and costs are reverse. Its main benefit is its intuitive plausibility: a tropist need not posit the mysterious entity which can be wholly present at many places at once. The ontological parsimony might be counted as one more benefit: it does not posit universals. Its main cost, on the other hand, is that it seems difficult to provide easy solutions to the problems stated above – the one-over-many problem, the similarity relation problem, and the Hume's question.

As Heil sees it, the benefit of a universalist theory is not so large as it seems to be at first glance. As to the first and the second problem, a proponent of tropes can freely speak of objects 'sharing' properties (concerning the one-over-many problem) or of distinct objects possessing 'the same' property (concerning the similarity relation). It is true that in these cases, 'same' means not self-sameness (strict identity) but exact similarity. However, as Heil sees it, 'this is the sense of 'same' intended by non-philosophers when they speak of distinct objects possessing the same characteristic'[56].

As to the third problem, the induction problem, the benefit of a universalist theory is not so large either. If we build causal powers into the properties (whether they

---

have by and large regarded objects as made up of bundles of tropes' and this he rejects (cf. ibid.). I will, however, use more popular term 'tropes' in this thesis.

[54] See Chapter 13, 'Modes', of *FOPV*.

[55] *FOPV*, p. 137.

[56] Ibid. p. 139.

are universals or tropes), we do not have to worry about whether objects possessing similar properties behave similarly or not. Properties are similar only if their contribution to the dispositionalities of their possessors is similar. It is true that objects could behave differently in the future because they cease to possess properties similar to these, but this point applies equally to the universalist theory as well. Identity is doing no work here.[57]

I do not make the final decision about the debates between the universalist theory and the trope theory: it should be the topic of another book. It suffices, however, to notice that the trope theory is at least as prospective as the universalist theory. As I see it, both theories have their advantages and disadvantages, and we should consider them as a kind of research programmes. I myself will take the trope theory. The assessment of the theory should be made on the basis of the whole plausibility – what problems it solves and what problems it leaves unsolved. I will construct my view on the basis of trope theory, in the next section.

6 - 3 A Brief Look at a Causal Trope Theory

In the following, I will expound my own view on properties, a Causal Trope Theory. The theory has three components, which are listed as follows:

Sparseness: Properties are *sparse*. That is to say, not every predicate corresponds to a genuine property.

Tropism: Properties are tropes.

Essentialism: Properties have their own causal profiles essentially and, indeed, are reduced to the clusters of causal powers.

It should be quite easy to see the similarity between **Sparseness** and Heil's denial of the principle (φ) (and his denial of the levels conception of reality).[58] Also, **Tropism** shows

[57] See *FOPV*, pp.143-145.
[58] David Armstrong and David Lewis are the main philosophers who advocate the

that I share with Heil a tropist intuition on properties: I am a robust particularist and I deny universals. As to **Essentialism**, I depart from Heil's view, and this is the topic of the sub-sections below.

6-3-1 Three Views on the Relation between Properties and Causal Powers

We have an intuition that there is *some* connection between properties and causal powers. This intuition is obvious when we consider the properties of fundamental physics. Physicists construct theories and make experiments based on the theories, trying to capture the real features of the world. When they construct a theory and make experiments, positing the fundamental physical properties such as charge of electrons, charm of quarks, and so on, it is crucial how those fundamental physical properties causally behave.

But what exactly is this connection? And how strong is this connection? There are two views on the connection between properties and their causal powers – Humean Theory and Causal Theory:

> Humean Theory[59]: The connection between properties and their causal powers is *contingent*.
> Causal Theory[60]: The connection between properties and their causal powers is *necessary*.

According to Humean Theory, we can identify properties independent of their powers. A property can have different causal powers in a possible world where different natural laws obtain. Which natural laws obtain in which possible worlds is a contingent matter according to Humean Theory, and causal powers of properties depend on natural laws.

'sparse theory' of properties. However, they are not trope theorists as me. Armstrong is a universalist and Lewis is a class nominalist. See, for example, Armstrong, *What is a Law of Nature?* (Cambridge: Cambridge University Press, 1983) and Lewis, *On the Plurality of Worlds* (Oxford: Blackwell, 1986) for their views on properties.

[59] See, for example, Armstrong, *What is a Law of Nature?* (Cambridge: Cambridge University Press, 1983) and Lewis, *On the Plurality of Worlds* (Oxford: Blackwell, 1986).

[60] See, for example, Shoemaker (1980), Shoemaker (1998), Martin (1980), Hawthorne (2001), Heil (2003), Bird (2007).

Therefore, the connection between properties and their causal powers is also contingent. As Causal Theory has it, on the other hand, there is a stronger connection between properties and their causal powers. It is impossible for a property to behave differently in a possible world.

I make two comments here. First, notice that 'causal powers' are sometimes expressed as 'causal behaviours', 'dispositions', 'causal features', or 'causal profiles'; I don't distinguish them. Second, although I sometimes use the expressions such as 'a property behaves this way or that way' or 'a property has such and such causal powers', it should be understood as abbreviated expressions. What behave or have causal powers are not properties themselves but the *possessors* of properties. If an electron with a negative charge is attracted by a metallic plate with a positive charge, then the behaviour – the moving towards the plate – is the behaviour of the electron, not of the positive charge of the electron. The electron behaves in that way in virtue of having the charge.

Now, we can further distinguish the two versions of Causal Theory – Weak Causal Theory and Strong Causal Theory[61]:

> Weak Causal Theory: For every property, there is a causal profile the possession of which is a *necessary* condition for having that property.
> Strong Causal Theory: For every property, there is a causal profile the possession of which is *necessary and sufficient* condition for having the property.

Both versions insist that one and the same property *cannot* have different causal profiles. That is to say, both versions of causal theory insist that causal profiles *supervene* on properties. Weak Causal Theory stops here: it allows the possibility that two different properties have the same causal profile. Strong Causal Theory, on the other hand, goes further: it insists that properties and causal profiles are in one-to-one correspondence.

It is fairly natural to interpret Strong Causal Theory as the theory insisting that we need not add anything to causal profiles in order to get properties, the theory insisting that properties are entirely exhausted by causal profiles or that properties *are* causal profiles. We may call it the *reductive* causal theory: it insists that properties are

[61] Cf. J. Hawthorne, 2001, 'Causal Structuralism', *Philosophical Perspectives* 15.

reduced to causal profiles. However, a reductive theory is not the only interpretation of Strong Causal Theory. We might be able to say that causal profiles and properties are different entities although having a causal profile is necessary and sufficient for having a corresponding property. Sydney Shoemaker, one of the main advocates of causal theory of properties, once expresses his view as follows:

> …properties are causal powers. (Shoemaker 1980, p. 210)

In a recent paper, however, he uses a slightly different expression:

> …the properties…have [causal features] essentially, and are individuated in terms of them. (Shoemaker 1998, p. 413)

As I see it, the difference between two expressions reflects the difference between the reductive reading of Strong Causal Theory and the non-reductive reading of it. In any way, we should notice that both versions of Shoemaker's formulation are Strong Causal Theory.

Now it should be clear that the third component of the Causal Trope Theory, **Essentialism**, indicates that it can be classified as the strongest causal theory, the reductive Strong Causal Theory. It explicitly says that properties are nothing over and above the cluster of causal powers.[62]

6-3-2 What the Causal Theory of Properties Is Not?

Now I would like to make two comments on what the causal theory of properties (whether or not it is combined with trope theory) is *not*. First, any version of the causal theory of properties is not intended to apply to *all* properties. The causal theory applies to sparse properties which appear in genuine changes or causal explanations. Properties appearing in the ultimate physical theory are the typical examples of such properties. We might be able to extend the scope of the causal theory as to include the properties

[62] The proponents of the strongest causal theory include, for example, John Hawthorn and Alexander Bird. See Hawthorne (2006) and Bird (2007).

appearing in the folk physics or folk psychology. However, the mathematical properties , the logical properties, or properties such as 'being grue' are out of the scope of the causal theory. The causal theory is not intended to apply to these properties.

Second, the causal theory of properties is intended to answer the question: in virtue of what does a property F is the property F (instead of the other property G)? That is to say, it gives the criterion of identity within the category of property. Therefore, the causal theory of property itself is perfectly neutral to the debates between universalists and tropists. The causal theory itself does not tell us if we should regard properties as universals, classes of tropes, or classes of possible particulars. Shoemaker, for example, who is a significant defender of the causal theory, seems to be a universalist, while I am a tropist. The difference depends on how one should think of the concept of *causation*.

6-3-3 Compared with Heil's Identity Theory

In the remaining of the chapter, I will compare the Causal Trope Theory with Heil's view. First, I examine his identity theory of dispositions and qualities in this subsection. Then I will examine the intrinsicness of properties in the next subsection.

Let us look back to section 6-1, where I listed Heil's main ontological claims. The claim (4) says that all properties are both dispositional and qualitative at the same time. Heil calls this claim the Identity Theory. The identity theory is a version of neutral monism[63], where one and the same property is dispositional and categorical ('qualitative' in Heil's terminology) at the same time.[64] We can describe a property both as dispositional and as qualitative. Heil says:

> Being spherical is a manifest quality of a snowball. But it is in virtue of being spherical that a snowball could, for instance, roll: sphericity is, it would seem, a power possessed by the snowball. … the snowball's sphericity *is* a quality possessed

[63] Stephen Mumford presents a very similar view, which can be also classified as neutral monism. See Mumford, 1998, *Dispositions*. I will discuss Mumford's view in the next chapter.

[64] Heil calls his own view 'the identity theory'. See Chapter 11, 'The Identity Theory' of *FOPV*.

by the snowball and *is* a power.[65]

It is clear enough that Heil's identity theory is not Humean Theory. It is also clear enough that Heil's identity theory is not Weak Causal Theory. If qualities are identical with powers, then causal powers should be necessary and sufficient for qualities. (I will argue against Weak Causal Theory later in this chapter.) Heil, however, does not take the strongest Causal Theory, the reductive Strong Causal Theory. He says:

> Think of a quality: being white, for instance […]. It is surely in virtue of its being white that a cupcake would look white […]. Being white and being sweet are powers of the cupcake to affect […]. The mistake […] would be to conclude from this that whiteness and sweetness are *mere* powers.[66]

Although Heil thinks that qualities are powers, he denies the view that qualities are nothing over and above powers.

What exactly, according to Heil, is added to powers? Why does he need qualities in addition to dispositions at all? What makes him believe that dispositions are qualities as well? We can at least insist that it is quite difficult to imagine the identity between dispositionality and qualitativity. While dispositions are essentially *for something*, qualities are not. Two descriptions, dispositional and qualitative, are apparently *in*consistent with each other. How can two descriptions that are so radically different from each other be applied to one property? Is it the same case as 'heat = mean kinetic energy' or 'being in pain = C-fibre excitation'? Let us see the arguments Heil presents for the identity theory. Heil points out our ability of abstraction ('partial consideration'). We can describe one and the same property both as quality and disposition. Why can we do this? That is because, as Heil sees it, we have the ability of abstraction. Just as we can think about a man's height without thinking about his skin colour, so can we think about dispositionality of a property without thinking about its quality.[67] Heil also mentions a Necker cube (i.e. which has recourse to Gestalt perception) as an example. We can perceive a Necker cube one way or another.

---

[65] *FOPV*, p.112.
[66] *FOPV*, p.113.
[67] See *FOPV*, pp.118-120.

However, we don's say that the cube has two distinct properties with respect to each perception.

Here, Heil seems to think about properties parallel to objects (substances). But I think there is an asymmetry between objects (substances) and properties. One and the same object (substance) can be described in many ways. We can think that the object has many properties with respect to some of the descriptions. The reason why this is possible is that an object has substratum (bare particular) as its component, a component which is perfectly neutral to the ways we describe the object. On the other hand, properties are ways objects are.[68] So it seems that the ultimate description (if there is such) of a property should exactly correspond to the feature of the property. This asymmetry seems to cause some uneasiness as to treating properties and objects in the same way. I will return to this problem when I discuss the problem of quiddity in Chapter 7.

Heil points out that the world of pure powers faces a great difficulty. The world of pure powers (dispositions) is, as he sees it, just like the world of dominos where all there is to each domino is to be toppled by the former domino and to topple the next domino – nothing occurs without quality[69]. As he sees it, qualities are needed to stop this regress. Introducing qualities and identifying them with dispositions (powers), we can avoid the difficulty.

What exactly about the quality stops the regress in the pure powers world? It might be the *actual* occurrence of dispositions. But how can we identify the occurrence with the possibility? They seem to be totally different with each other. Heil never offer an explanation on this problem. I am not quite sure if Heil's argument works well, but so far as we can regard pure dispositions in themselves as intrinsic and actual, there seem to be no threat of infinite regress here. I will return to this problem in Chapter 7.[70]

---

[68] At least, this is what Heil himself takes for the theory of properties. See *FOPV*, p.12.

[69] See Chapter 10 of *FOPV*.

[70] The difficulty in identifying a power with a quality may have some connection with the difficulty in identifying a potentiality with the exercise of it. It might be suggestive to see E. J. Lowe's view on dispositions. According to his four-category ontology, the distinction between the dispositional and the occurrent (categorical) lies in the difference between the ways of characterizing substances. When we say 'This lump of gold has such and such micro-structure', we are characterizing it as possessing a mode (trope) of micro-structure which belongs to a non-substantial universal – a micro-structure. When we say 'This lump of gold has ductility', we are characterizing it

There is one more motivation for Heil's identity theory. Heil says:

> To my mind, the identity theory is independently attractive, but, even if it were not, it appears to win by default! Purely qualitative properties lack appeal, as do pure powers. Mixing these does not help matters, nor does turning dispositionality and qualitativity into aspects or kinds of higher-order property.[71]

This quotation, as I see it, best expresses what makes Heil take the identity theory. Heil seems to propose the identity theory as what can overcome the difficulties that pure quality theory and pure power theory encounter, *leaving the unintelligibility of the identity aside*. If he succeeds in reducing the unintelligibility well enough, we may accept this identity theory. However, as I have seen it so far, he does not succeed in doing it well enough.

6-3-4 Intrinsicness of Properties

According to Heil, all dispositional properties (not to mention qualitative properties) are

---

as belonging to a kind gold which possesses a non-substantial universal – ductility. The former corresponds to the occurrent (categorical, qualitative) predication and the latter to the dispositional predication. In this example, two different non-substantial universals (micro-structure and ductility) were used to explain the dispositional/occurrent distinction. When we characterized a particular lump of gold *dispositionally*, we mentioned ductility. When we characterized it *occurrently*, we mentioned micro-structure. Can we characterize it dispositionally, mentioning micro-structure? The answer is probably yes. We have only to admit that a kind can possess a non-substantial universal. Can we, then, characterize a lump of gold occurrently, mentioning ductility? The answer is probably no. I believe that we can't characterize a lump of gold occurrently by means of ductility. For, what is a particularized ductility? It's the very thing we should call a *manifestation* of ductility! This will explain the unintuitiveness of Heil's identity theory. According to their ontology, universals such as solubility can be particularized. I believe, however, the idea of a particularized solubility conflicts with the concept of solubility as a disposition. We might be able to say that there are two sorts of non-substantial universals in Lowe's ontology. As to the universals such as micro-structure, geometrical features, and so on (which roughly correspond to so-called categorical properties), we can use them to characterize particular objects both dispositionally and occurrently. As to the universals such as solubility, ductility, and so on, we can only use them when we characterize particular objects dispositionally. Cf. Lowe (2006), especially Chapter 8.

[71] *FOPV*, p.120.

intrinsic (see claim (5) in section 6-1). But here we should note that this intrinsicness of the dispositional is a consequence of identifying the dispositional with the qualitative. In Heil's ontology, dispositional conception, in itself, does not include intrinsicality. It is, as it were, *derived* from intrinsicality of the qualitative. In contrast to this, I attach intrinsicality to the dispositional independent of the qualitative. I will return to this matter in the next chapter.

The philosophers who contribute to the dispositions debates do not agree with the meaning of 'dispositions', which causes some confusion. Let us see how the dispositional/categorical distinctions are characterized by the main contributors. According to Lowe's four-category ontology, the dispositinal/occurent (categorical) distinction lies in the difference between the ways of characterizing substances. When we say 'Mr. Jones is short-tempered.', we are characterizing him as belonging to a *kind* which possesses a non-substantial *universal* – short-temperedness. When a disposition is ascribed to a particular object, the particular object does not really possess the disposition. It just belongs to a kind which possesses the disposition. On the other hand, when we say 'Mr. Jones becomes angry', we are characterizing him as possessing a *mode* – being angry. In this case (unlike fragile vase case), he does not lose short-temperedness (disposition) when he becomes angry. We should notice that in Lowe's ontology, the way the dispositional exist is totally different from the way the occurrent exist. The dispositional are potentialities, while the occurrent are exercises of potentialities.[72]

Let us see the following view stated by Stephen Mumford:

> Disposition ascriptions are categorical in the sense that to say that something has a dispositional property is to say that something has a property *actually*. (1998, p.37)

If we call those who agree with this view 'disposition actualists', we can include George Molnar in this 'disposition actualists' as a core member (cf. Molnar 2003, Chapter 5). Being both a pure dispositionalist and a disposition actualist, Molnar's

[72] For E. J. Lowe's concept of dispositionality based on his 'Four-Category ontology', see, for example, his *The Four-Category Ontology: A Metaphysical Foundation for Natural Science*, Chapter 8.

intuition about dispositions can be sharply contrasted with E. J. Lowe's intuition.[73]

Is John Heil a disposition actualist? Yes. But we should notice how he takes a disposition actualist view. He takes a disposition actualist view just by identifying dispositions with actual qualities. Remember his argument from the never-toppled dominos example. We can interpret this argument as an attack on non-actuality of pure powers. The world of pure dispositions, he argues, is just like the world of dominos where all there is to each dominos is to be toppled by the former domino and to topple the next domino – nothing occurs! He introduces the qualities and identifies them with dispositions (powers) to avoid the difficulty. But why can't he just say pure powers are actual in its own right without identifying them with actual qualities? This is because he shares with Lowe the intuition about dispositions (powers). Just identifying dispositions with qualities to avoid difficulty is not good way to solve the problem but to put off dealing with the problem.

All in all, I think that Heil's identity theory is unstable. You might take the disposition NON-actualist view (like Lowe and Armstrong). You might take the view that says there exist pure dispositions and they have full-fledged actuality (like me). But we cannot easily take the middle way just by presupposing the identity between dispositions and qualities.[74]

So far, I have just sketched the general idea of my own theory of properties, the Causal Trope Theory, comparing it with some other views (especially with John Heil's).

---

[73] Molnar also admits pure categorical properties ('non-powers' as he calls them). So he is, to be exact, a property *dualist*. See Molnar (2003), pp.158-172.

[74] There is one more worry about Heil's view. Some philosophers have regarded intentionality as the mark of the mental – 'Brentano's thesis'. What kind of account does Heil give to intentionality? He proposes that we tie the intentional character of states of mind to their dispositionality (cf. *FOPV*, p.210). 'Dispositionality underlies the projective character of thought.' (ibid.). Heil presents the internalist view about the content of thought on the basis of the intrinsic theory of dispositions. One worry remains. Heil's ontology apparently opposes Brentano's thesis. What criterion, then, can Heil give to distinguish the mental from the physical? For example, George Molnar, who is a property dualist, seems to regard quality as the mark of the mental (cf. Molnar (2003)). Without an adequate criterion of the mental, Heil might be led to panpsychism.

I have not defended the Causal Trope Theory, with contrast with a main rival, Humean Theory. This is the topic of the next chapter.

7 The Need for the Ontology of Properties (2): Defending Dispositionalism and Attacking Categoricalism

In this chapter, I will do two things: *defending* dispositionalism and *attacking* categoricalim.

I will first examine a typical and influential categoricalist view on properties which is pressed forward by Prior, Pargetter, and Jackson many years ago. I will then focus on an assumption of their argument, the Causal Thesis. The Causal Thesis has its origin in David Armstrong's argument against Gilbert Ryle's phenomenalism. I will examine Armstrong's argument and conclude that Prior, Pargetter, and Jackson's argument does not succeed in knocking down dispositionalism. After that, I will examine a recent argument by Bradley Rives which tries to defend Prior, Pargetter, and Jackson's view. I will show that Rives' argument does not succeed.

After that, I will try attack categoricalism, posing some serious problems for it.

7 - 1 Prior, Pargetter, and Jackson on Dispositions

In their seminal paper, 'Three Theses about Dispositions', Prior, Pargetter, and Jackson (hereafter, PPJ) argue that dispositions are *distinct* from their causal bases and that dispositions are causally *impotent*[75]. Their argument is considered to be one of the first attempts to argue for a categoricalist view on properties. I will start my support for a dispositionalist view on properties, with a close examination on their argument. What I will try to do in the following is to analyse their argument and to clarify what is the ontological assumption of their argument.

7 - 1 - 1 Three Theses (Overview)

In 'Three Theses about Dispositions', PPJ argue for the following three theses about

[75] Prior, Pargetter, and Jackson, 1982, 'Three Theses about Dispositions'

dispositions[76]:

The Causal Thesis: It is a necessary truth that dispositions have causal bases.

The Distinctness Thesis: These bases are distinct from their attendant dispositions.

The Impotence Thesis: Dispositions are causally impotent with respect to their manifestations.

Let us see what the theses mean and what is the outline of their argument, just briefly (I will closely examine the argument in the next section).

The Causal Thesis says that every disposition must have its causal basis. What is 'a causal basis' of a disposition? By 'a causal basis', they mean '*the property or property-complex* of the object that, together with ... the antecedent circumstances[,] is *the causally operative sufficient condition* for the manifestation in the case of 'surefire' dispositions, and in the case of probabilistic dispositions is causally sufficient for the relevant chance of the manifestation'[77]. For a disposition, e.g. fragility, we can specify a pair <knocking, breaking>, where the former (knocking) indicates a stimulation or a triggering cause and the latter (breaking) indicates a manifestation. To ascribe a disposition to an object is to say that the object satisfies the pair that we specify for the disposition. To ascribe fragility to a glass is to insist that the glass satisfies the pair <knocking, breaking>: if it were knocked, it would break. The causal basis of the fragility is the property possessed by the glass which is, together with the knocking of it, causally operative sufficient condition for the breaking of it. As they see it, a property of having molecular bonding α is the causal basis in this case. The Causal Thesis says that for every disposition there necessarily exists a causal basis of it.

Now PPJ ask one question: what is the relation between the disposition, e.g. fragility of a glass, and its causal basis, e.g. the molecular bonding α of the glass? Are they identical or distinct? The Distinctness Thesis says that they are distinct: the fragility of the glass is not its molecular bonding α. The main reason they offer for the Distinctness Thesis is so-called 'multiple realizability of dispositions': a glass with a slightly different molecular bonding β than α would be fragile, as well as a glass with

[76] ibid., p.251.
[77] ibid. Italicized by me.

the molecular bonding α would. Moreover, something that has a very different microstructure than a glass with a molecular bonding α, e.g. a vase with microstructure γ could be fragile as well. The fact that one and the same disposition may have different causal bases suggests that the disposition cannot identical to *all* of them; otherwise the transitivity of identity is violated. If α is identical with fragility and fragility is identical with β, then α is identical with β (transitivity of identity), whereas, *ex hypothesi*, α is not identical with β.

Suppose a glass is knocked and breaks. Is its fragility the cause (or a part of the cause) of its breaking? The Impotence Thesis says that it is not. It insists that every disposition, including fragility of course, is causally impotent: they do not play any role in bringing about the effects. PPJ argue that the Impotence Thesis is implied by both the Causal Thesis and the Distinctness Thesis. Here is the argument:

> By the Causal Thesis, any disposition (and thus fragility) must have a causal basis. This causal basis is a sufficient causal explanation of the breaking *as far as the properties of the object are concerned*. But then there is nothing left for any other properties of the object to do. By the Distinctness Thesis the disposition is one of these *other* properties, ergo the disposition does nothing.[78]

Suppose a disposition manifests itself. We are asked what caused the manifestation. The natural answer should be this: it is the basis of the disposition that caused the manifestation event. But if dispositions are distinct from their bases, we cannot say that the disposition itself caused, or at least causally contributed to the occurrence of, the manifestation.

While the argument seems, at least at a glance, plausible, the conclusion, the Impotence Thesis, is rather surprising. If dispositions do nothing in bringing about their manifestation, why do we have to posit dispositions in the first place? The Impotence Thesis has it that every disposition does nothing in bringing about its manifestation. If this is true, why we need dispositions as ontological items? If they are causally impotent, we wouldn't have any access to them. Therefore, we cannot know that there really are such things in the world at all, can we? In the following, I will examine the argument

[78] 'Three Thesis about Dispositions', p.255

and make it clear what Prior, Pargetter, and Jackson's argument shows and what it does not show.

7 - 1 - 2 Prior, Pargetter, and Jackson's Argument for the Causal Thesis

The Causal Thesis has it that every disposition necessarily has causal basis. The whole of the first section of their paper is devoted to argue for the Causal Thesis. As I mentioned in the previous section, they define the causal basis as follows:

> A causal basis is the property that, together with the antecedent circumstances, is the causally operative sufficient condition for the manifestation.

It is not quite clear, though, exactly what the Causal Thesis means and what is the causal basis of dispositions. Ask the following question: what is it like for the Causal Thesis to *fail* to obtain? The Causal Thesis appears, at least at a glance, so trivial that it is difficult to imagine the situation where the thesis fails to obtain. Keeping this point in mind, let us see their argument for the Causal Thesis.

PPJ's argument for the Causal Thesis goes as follows. Let us consider two cases – we might live in a deterministic world or in an indeterministic world. Let us start with a deterministic case (that is, we live in a deterministic world). Suppose that a glass is knocked at t in a possible world that is the closest to ours. The same laws as ours obtain in this possible world and the possible world is deterministic because our own actual world is deterministic. Either the glass breaks at $t+\delta$ or not, in this possible world. If it does not break, then the glass is not fragile. If it breaks, then the glass is fragile and 'there will be a causally sufficient antecedent condition operative in producing the breaking – that follows from Determinism'[79]. Therefore, there is no counterexample to the Causal Thesis in the deterministic case. However, there *seem to be a putative counterexample* for the Causal Thesis in an indeterministic case. Suppose that we live in an indeterministic world and that two rubber bands A and B are fully examined and proved to be causally alike with each other in this world. Suppose, also, that A and B are stretched at time t and that A returns to its original length at time $t+\delta$, while B does not.

[79] 'Three Theses about Dispositions', p.252

It seems to follow that *A* has a disposition, elasticity, and *B* lacks it, while, *ex hypothesi*, *A* and *B* are causally alike. Therefore, this seems to be the case where *A* has a disposition (elasticity) that, contrary to the Causal Thesis, doesn't have a proper causal basis. This could happen, so insists the defender of the putative counterexample, in any indeterministic world. But Prior, Pargetter, and Jackson asks: could it happen indeed? If *A* and *B* are causally alike in an indeterministic world, it should mean that *A*'s and *B*'s stochastic behaviour, their probability of returning to their original lengths, are the same. Therefore, even if *A* returns to its original length and *B* doesn't at this very moment, it doesn't follow that they have different dispositions. The only possible counterexample being rejected, the Causal Thesis holds – so they argue.

Let us extrapolate their argument. Suppose that every physical event is causally determined, as determinism says it is. Every physical event, including manifestations of dispositions, is determined by the previous physical states, or the part of the physical states. So, for every disposition there is a proper physical state that is causally sufficient for bringing about its manifestation. So far, so good. Suppose, then, some physical events are not causally determined as determinism insists. That is to say, we examined both rubber bands, *A* and *B*, and found no difference between them; nevertheless *A* and *B* behave differently. What should we make of this case? It is most likely that we didn't examined *A* and *B* precisely enough. If we examine them more carefully and more precisely, perhaps we will find the difference between them. With those fine-grained properties, we will be able to predict, deterministically, their behaviours. But what if no difference in physical properties was found, however carefully we examined the objects? We will, then, take it that the behaviours of *A* and *B* are probabilistically, thought not deterministically, determined. Now, how should we specify their probabilities? We must specify an ensemble that is constituted by the members with the same physical property and see how many of the members behave such and such way. But this ensemble include both *A* and *B* as its member, because *ex hypothesi* they are the same in respect of physical properties. Therefore, we *never can* ascribe the different probabilities to *A* and *B*. Otherwise, we simply have to give up scientific investigations and leave the world completely un-understandable. This is, as I extrapolate it, the general line of their argument.[80]

[80] Cf. Ibid.

I don't object the inferences in this argument, nor do I object the conclusion of it. I must ask, however, why does this argument matter to dispositions at all? Does this argument illuminate some aspects of dispositions? It seems that this argument just reconfirms the general process of how we conduct scientific investigations. As I see it, Prior, Pargetter, and Jackson's alleged argument for the Causal Thesis *does not prove* the necessary existence of the causal basis of dispositions at all. It just *assumes* both that every physical events are, deterministically or indeterministically, determined by the previous physical states of the world and that dispositions *supervene* on those physical states. The latter assumption should be noticed when they insist that the only possible counterexample of the Causal Thesis is the case in which '*A* and *B* are causally alike but differ in their chance of returning to their original length the next time they are stretched.'[81]

However, they are not to be blamed for not offering an effective argument to prove the Causal Thesis. In fact, in recent studies on dispositions, a causal basis of a disposition is sometimes simply *assumed* (not *proved* to exist).[82] Furthermore, David Lewis *introduced* the intrinsic causal bases to deal with a counterexample to the conditional analysis of dispositions, that is, 'finkish' dispositions[83]. It is also worth mentioning that the Causal Thesis does not specify the nature of the causal basis, but just has it that a causal basis of a disposition is the proper part of the physical states of the world that determines the manifestation of the disposition. The thesis, therefore, does not preclude the possibility that the causal basis of a disposition is *itself* a disposition as well. We will return to this problem later in this chapter.

I do not bring in a final verdict upon Prior, Pargetter, and Jackson's argument in this stage. In the next section, I will focus on David Armstrong's view on dispositions, for the Causal Thesis of PPJ has its origin in Armstrong's requirement for causal bases. After I further examine the *nature* of causal bases and the *relation* between dispositions and their causal bases, I will say something conclusive about Prior, Pargetter, and Jackson's argument.

---

[81] 'Three Theses about Dispositions', p.252

[82] See, for example, Lewis (1997), Mumford (1998) etc.

[83] See C. B. Martin, 1994, 'Dispositions and Conditionals', *Philosophical Quarterly* 44, pp.1-8.

7-2 The Nature of Causal Bases and the Relation between Dispositions and Their Causal Bases

The Causal Thesis of Prior, Pargetter, and Jackson has its origin in David Armstrong's work. In this section, I will first examine David Armstrong's argument for the categorical bases, where Armstrong attacks Gilbert Ryle's anti-realistic view on dispositions and goes on with what he calls 'a Realist view'. I will argue that Armstrong's attack on Ryle is generally right, but that his requirement for bases of dispositions needs amendments. I will, then, offer a more satisfactory formulation of the requirement for bases (7-2-2). This new formulation will help to dissipate some confusion in the debates on dispositional bases and the possibility of bare dispositions.

7-2-1 Ryle and Armstrong on Causal Bases

In *A Materialist Theory of Mind*, D. M. Armstrong argues against Ryle's view which Armstrong calls the *Phenomenalist* or *Operationalist* account of dispositions. He then argues for his own view which he calls the *Realist* account of dispositions. Armstrong characterizes the Realist view by means of a principle which is very close to the Causal Thesis of PPJ. In the following, I will examine Armstrong's argument. That would help to clarify what the Realist view on dispositions says about dispositions, and therefore, what exactly PPJ's Causal Thesis means.

Ryle's view on dispositions is shown in the following quotation:

> To possess a dispositional property is not to be in a particular state, or to undergo a particular change; it is to be bound or liable to be in a particular state, or to undergo a particular change, when a particular condition is realized.[84]

To this, Armstrong opposes what he calls 'the Realist view':

> According to the Realist view, to speak of an object's having a dispositional property entails that the object is in some non-dispositional state or that it has some property

[84] G. Ryle, *The Concept of Mind*, p.43

(there exists a 'categorical basis') which is responsible for the object manifesting certain behaviour in certain circumstances, manifestations whose nature makes the dispositional property the particular dispositional property it is.[85]

It is clear that the Realist view requires the necessary existence of the causal bases for all dispositions; the requirement is quite similar to that of the Causal Thesis of PPJ. Let us list both theses. The Causal Thesis of PPJ is as follows:

The Causal Thesis: It is necessary that every disposition has a causal basis (property) which is the causally operative sufficient condition for the manifestation.

And Armstrong's requirement for bases (**ARB**) is simply expressed as follows:

ARB: It is necessary that every disposition has a non-dispositional (or categorical) property which is responsible for its manifestation under suitable circumstances.

One obvious difference between them is that **ARB** (that is, the Realist view) explicitly requires *non-dispositional* property as a causal basis, where the Causal Thesis of PPJ remains neutral in whether a basis should be non-dispositional or not.

Armstrong presents what he calls an *a priori* argument to support the Realist account of dispositions. The argument has five steps. Let us see each step one by one.

Step 1. Suppose that we have a rubber band that has manifested a disposition, elasticity, on past occasions when the conditions of manifestation obtained: it stretched one inch whenever force *F* was applied to it. Suppose the rubber band is not in the condition of manifestation now. 'Now one essential thing about dispositions is that we can attribute them to objects even at times when ... [the conditions of manifestation] do not obtain.'[86] We, therefore, should be able to say, of the band, that if it were under force *F* at T_1, a time when it is not under force *F* in the actual world, it would stretch one inch. Now Armstrong asks one question: what warrant do we have for that counterfactual statement? The Realist has an answer. The realist has good reasons to

[85] D. M. Armstrong, *A Materialist Theory of Mind*, p.86
[86] ibid.

believe 'that the categorical state of the band which is responsible for its stretching one inch under force *F* obtains at T_1'[87]; this belief has been confirmed by the past behaviours of the band. Given that the band has the categorical state at T_1, it must stretch one inch under force *F*.

Step 2. But what answer, challenges Armstrong, can the Phenomenalist give? The only answer that Armstrong contrives for the Phenomenalist is that the fact that 'numerically the same' band stretched one inch under force *F* on past occasions allows us to infer the present and future behaviour of this band. But this is not a satisfying answer because the band might change its dispositional property over a period of time.

Step 3. Armstrong considers a conceivable reply, by the Phenomenalist, to the objection. '[The Phenomenalist] may reply "We have every reason to think that the relevant categorical properties of the object are unchanged at T_1, so we have every reason to think that the dispositional properties are unchanged."'[88] But this reply, as Armstrong sees it, assumes that there is some connection between categorical properties and dispositions. The Phenomenalist should not be allowed to rely on any connection between categorical properties and dispositions.

Step 4. The Phenomenalist might still insist, so Armstrong suggests, that a *contingent* connection between a categorical property and a disposition is available.

> ...[S]ince he has asserted that the connection between 'categorical basis' and dispositional property is not a necessary one, he can only be arguing that there is a contingent connection between categorical properties and the fact that the band has that dispositional property at T_1.[89]

But how could we establish this kind of connection, a contingent connection between properties and *un-manifested* dispositions? There could be a contingent connection between a property (say, a property of a die at *t*) and a *manifested* result (say, the die is rolled and the result is '3' at $t+\delta$), if there could be fundamentally indeterministic processes. But we cannot establish, so insists Armstrong, a contingent connection between a property and an un-manifested disposition itself.

---

[87] ibid.
[88] ibid. p.87
[89] ibid.

Step 5. Barring all the possible paths the Phenomenalist could take to authorize disposition ascriptions, Armstrong concludes that the Phenomenalist account of dispositions is not tenable. To deny the necessary connection between dispositions and causal bases leads to allow the possibility that two objects that are exactly similar in their categorical properties differ in their dispositions, which amounts to deny the Principle of Sufficient Reason: if the Phenomenalist were right, the world would resist any attempt to understand it and would be completely un-understandable.

In response to this argument, D. H. Mellor points out that if it were sound, then it would have the consequence that contradicts Armstrong's earlier insistence on the contingency of a categorical basis of a disposition.[90] According to the above argument, there cannot exist a contingent connection between categorical properties and dispositions; there is either a necessary connection or no connection at all. As his earlier and later works show, however, Armstrong insists that natural laws obtain contingently.[91] This means that a categorical property behaves this way in this world, while it might behave another way in another possible world where different laws obtain. As Mellor sees it, the contradiction comes from the fact that Armstrong, in the above argument, tacitly and mistakenly infers from the necessary existence of a connection (between categorical properties and dispositions) to the existence of a necessary connection.[92] Mellor points out that the original requirement, the requirement before the erroneous inference, of Armstrong's Realist view should be the necessary existence of a (contingent) connection.

Michael Fara, following Mellor's argument, takes it that what is insisted in Armstrong's Realist view, before the erroneous inference, should be that every dispositions have *some* categorical basis, not that every dispositions have the *particular* categorical basis.[93] He, then, insists that Armstrong's view is no more enough to preclude the possibility that an object might change dispositionally without changing categorically than the Phenomenalist view. He also insists that this preclusion, although Armstrong makes an effort to do it, is not needed for authoring dispositional ascription

---

[90] D. H. Mellor, 1974, 'In Defence of Dispositions', p.165.
[91] See Armstrong, *What is a Law of Nature?*
[92] D. H. Mellor, 1974, 'In Defence of Dispositions', p.165.
[93] M. Fara, 'Dispositions', *Stanford Encyclopedia of philosophy*, http://plato.stanford.edu/entries/dispositions/

after all.

> We have no problem ... with supposing ... [the] objects don't change categorically while we're not looking at them. Why then should we anticipate an analogous problem ... with supposing that objects don't change dispositionally, whether we're looking at them or not (provided they haven't change categorically)?[94]

Fara concludes his review of Armstrong's *a prior* argument with the observation, which he draws from Mellor (1974), that Armstrong's problem is merely the problem of induction.

The situation is a bit complicated. Armstrong insists that, in order to authorize disposition ascriptions, we need some connection between a disposition and its categorical basis. But the connection, if there is any, must be a very strong one, a necessary connection, that is never available to the Phenomenalist; otherwise, there could be no connection at all. Mellor insists that there could be a middle way; he insists that a contingent connection is available. But, then, is the middle way available to the Phenomenalist as well? If so, does the purported distinction between the Realist view and the Phenomenalist view melt away?

As I see it, Armstrong is quite right in insisting that Ryle cannot authorize disposition ascriptions and insisting that it is a serious problem for him. Let us see, first, what is the point at issue between two parties, the Realist (Armstrong etc.) and the Phenomenalist (Ryle etc.). Both agree that there are ordinary things like glasses, vases, and so on. When a glass is hit at t and breaks at $t+\delta$, both agree that the hitting of the glass (an event or a process) exists at t and the breaking of the glass (another event or a process) exists at $t+\delta$. What about the supposed regularity that has obtained and will obtain concerning the fragile glass? I believe that Ryle, as well as the Realist, must concede the existence of the regularity in some way or another, because he takes 'the inference ticket view' on laws and dispositions. Whatever the nature of the world might be for Ryle, he must concede that the world is such that we can use the inference ticket, the ticket that allows us to infer from the hitting of the glass to the breaking of it. But, then, we need to know which glasses this ticket is for. Without this specification, we

[94] ibid.

cannot use the ticket. If, as the Phenomenalist would insist, the rubber band *is not in a particular state* when it is not under force *F*, how the Phenomenalist is able to distinguish the objects to which we can apply the ticket, from the objects to which we cannot apply the ticket. In the same way, we can ask how we issued the ticket in the first place without this specification. I believe this is what Armstrong has in mind. As I see it, when Armstrong points out, against the Phenomenalist who relies on the numerical identity, that the rubber band might change its dispositional property over a period of time, he misses the point. The possibility of an object's changing dispositionally over time is not the point at issue. The most serious problem for Ryle is, pace Mellor and Fara, not the problem of induction. I conclude that Ryle's view on dispositions is not consistent. He cannot *both* insist that to possess a dispositional property is not to be in a particular state *and* take the inference ticket view on laws and dispositions. Which is to be taken? Ryle must take the latter, the inference ticket view; otherwise, as Armstrong rightly points out, Ryle comes to reject the Principle of Sufficient Reason.

Armstrong is also right in insisting that, in order to authorize disposition ascriptions (or inference tickets) we need some *bases* or *grounds* for dispositions. We must, however, note that bases or grounds are needed, as far as Armstrong's argument shows, just for indicating to which objects we can apply an inference ticket. The following consideration might make this point clear. Why is Ryle inclined to reject a particular state of an object when dispositions are not manifested? I suspect that Ryle's rejection comes from his worries about the observability, or the epistemic integrity, of dispositions. We can certainly observe the hitting and the breaking of the glass. But we cannot observe the disposition itself. Therefore, Ryle might insist, when we ascribe a disposition to an object, we must not say that the object is in a particular state. But this reasoning is wrong. Suppose that the world consists of some macroscopic (middle-sized) objects and just four properties – being round, being triangular, being blue, and being red – which are all *observable* properties (at least we can plausibly suppose so). If we observe that all the round objects have changed their colour from blue to red (or from red to blue) and all the triangular objects have not changed their colour whenever they were hit, then we will say that a round object is disposed to change its colour and a triangular object is not. How we apply the inference ticket that

allows us to infer from ‘an object is hit’ to ‘the object will change its colour’? We apply this inference ticket to round objects, not triangular objects. In this case, the basis of the disposition, ‘changing colour’, is the shape (roundness) of the objects. Ryle does not need to worry about the observability at all. The roundness and triangularity (of macroscopic objects) are perfectly observable!

The considerations so far make it clear why we need bases for dispositions and what conditions these bases must satisfy. Although Armstrong is right in insisting the need for bases, he does not formulate the requirement for bases satisfactorily. First, it presupposes that the bases are ‘non-dispositional’ (or ‘categorical)’ properties. As I argued above, the bases of dispositions are required to pin down the objects that inference tickets are applied to and to issue, in the first place, an inference ticket. As far as this argument shows, the bases need not be non-dispositional or categorical: the argument does not rely on the nature of the bases. We can say quite rightly that a basis of a disposition might be a dispositional property. If Armstrong wants to deny this, he needs another argument.

The second and more important defect of Armstrong’s formulation is that, in his formulation, the connection between dispositions and bases are not strong enough. Armstrong’s original statement can be read such that every disposition has *some* basis, not that every disposition has the *particular* basis. But if this means just that for every disposition there exists a property in one-to-one correspondence, this does not preclude the possibility that an object might change dispositionally without changing categorically (to use Armstrong’s term). We do need a kind of necessity concerning the connection between dispositions and their bases. In my opinion, Armstrong should have formulated his requirement for dispositions’ bases by means of supervenience. This is the subject of the next section.

7-2-2 The Minimal Requirement for the Causal Bases

For the substitution of Armstrong’s formulation of the requirement for the causal bases, I recommend **the Minimal Requirement for the Bases of dispositions (MRB)**:

MRB: Dispositions supervene on bases, in that necessarily, for any disposition *D*, if

anything has *D* at *t*, there exists a basis *B* such that it has *B* at *t*, and necessarily anything that has *B* at a time has *D* at that time.[95]

I offer this formulation as a *minimal* requirement for disposition bases that Armstrong could accept. The possibility that an object might change dispositionally without changing its basis, the possibility which Armstrong worries, is precluded by the second 'necessarily' in the formulation. Note, also, that MRB says nothing about the nature of the bases itself. It just says that a basis is a *subvenient* property (or state) of a disposition. Whether we can ascribe a disposition meaningfully without a basis or not, is one problem. Whether a basis of a disposition is dispositional or non-dispositional is quite another problem. I will consider the second problem later. In this stage, however, I emphasize that it is quite important to keep these two problems separately.[96]

What about Mellor's worry? As Mellor sees it, if Armstrong's *a priori* argument were sound, it would establish far stronger connection, than Armstrong himself believes, between dispositions and bases: the necessary connection between dispositions and bases contradicts with contingency of natural laws which Armstrong explicitly insists on elsewhere. As I see it, Mellor confounds nomological necessity with metaphysical (or logical) necessity when he poses this worry. MRB would help us to make this point clear. The second 'necessarily' in the formulation can be interpreted in several ways according to the modal theory one prefers. If Armstrong interprets the second 'necessarily' as expressing nomological necessity (I believe that he should), then he can successfully preclude the possibility that an object change dispositionally without changing categorically (to use his term) and also hold the view that the way a property behaves is quite a contingent matter.

---

[95] This formulation is obviously based on Kim's formulation of 'strong supervenience'. See Kim, 1998, *Mind in a Physical World*, p.9.

[96] Let me compare MRB with Stephen Mumford's characterization of the basis. Mumford's formulation is as follows: the basis b, of any disposition d, is generally understood to be that property, or property-complex, in virtue which the object or substance has d. It is easy to see that this characterization is quite compatible with mine. It is generally understood that supervenience relation is less ontologically laden than 'in virtue of' relation. It is sometimes said that 'in virtue of' relation explains supervenience relation. The fact that MRB is less ontologically laden than Mumford's formulation might be thought of as a merit of my formulation compared with Mumford's, at least as a minimal requirement. Cf. Mumford (1998), p.97.

MRB helps to dispel Fara's worry, as well. Fara insists that Armstrong's view no more enough to preclude the possibility that an object might change dispositionally without changing categorically than the Phenomenalist view. This is not right. If the second 'necessarily' in MRB is interpreted as nomological necessity, an object cannot change dispositionally without changing categorically in any possible worlds where the same natural laws as ours obtain; this is, as I take it, what happens if we take Armstrong's view properly interpreted. If MRB is not satisfied, an object can change dispositionally without changing categorically even in our actual world; this is what would happen if we take the Phenomenalist view.

There is one more point to be mentioned about MRB. Compared with Prior, Pargetter, and Jackson's Causal Thesis and ARB, it is obviously abstemious about the specification of bases. As I have already mentioned, MRB does not specify whether a basis should be non-dispositional or not. What is more, it does not say whether a basis at *t* is causally operative or causally responsible for the manifestations. As I see it, the question of why bases are required for dispositions should be clearly separated from other questions – the question concerning the causal efficacy of bases, or the question whether we should reduce dispositions to causation or, conversely, causation to dispositions. Sometimes, a different question is conflated with the question of the requirement for bases. And this complicates many debates on dispositions. We will consider these other questions later.

7 - 3 Close Examination of the Distinctness Thesis

In the proceeding section, I formulated the minimal requirement for bases of dispositions. The minimal requirement, however, specifies very little about the nature of disposition bases; it just insists that dispositions *supervene* on basal properties. Formulated in such a sparing way, the requirement makes it clear that what is required and what is not required for a philosopher to be a disposition Realist (in Armstrong's and PPJ's sense). My point in the previous section was that we should clearly separate the question of what is the nature of bases from the question of whether bases must exist. Now, it's time to add something more to the minimum requirement. In this section, I will deal with the Distinctness Thesis of PPJ. The problem concerns a question: 'Are

dispositions distinct from their bases?'

Remember Prior, Pargetter, and Jackson's argument for the Distinctness Thesis: the thesis insisting that dispositions are *distinct* from their bases. The argument relies on the purported multiple realizability of dispositions by various causal bases. It goes as follows. There seems to be an empirical fact that a disposition may have different bases in different objects: a vase A with a molecular bonding α is fragile; a glass B with a crystalline structure β is also fragile. Suppose that the basis of A's fragility is α, while the basis of B's fragility is β. If dispositions are identical with their bases, then being fragile is identical with having molecular bonding α in A while being fragile is identical with having crystalline structure β in B. But we cannot hold both of these identifications at the same time because, from these identifications, it follows, due to the transitivity of identity, that having molecular bonding α is identical with having crystalline structure β, which is obviously false. Therefore, the argument concludes, dispositions are distinct from their bases. Let us call this argument *the argument from multiple realizability*.

7-3-1 Mumford's Reply: Token Identity Theory

In response to the argument from multiple realizability, Stephen Mumford insists that while the argument succeeds in denying type-type identity between dispositions and bases, it does not deny token-token identity between them. He tries to defend, against the argument from realizability, the property monism. According to the property monism, there is only one kind of properties (i.e. neutral properties) which is described both by dispositional predicate terms and by categorical predicate terms. He says:

> The property monist essentially wants to show that in saying *x* is D and *x* is C, where 'D' and 'C' are dispositional and categorical predicate terms respectively, we are not saying that there is some fact about *x* over and above instantiating C that makes it true that it is instantiating D.[97]

It should be clear, from this quotation, that Mumford insists that 'C' and 'D' denote one and the same property instance. If we could take token-token identity between

[97] S. Mumford, 1998, *Dispositions*, p.159.

dispositions and their causal bases, we would succeed in securing the causal potency of dispositions (disposition tokens, in this case). This is because, as Mumford sees it, the causal relation obtains not between universals in the abstract sense, but between property instances. He says:

> When we say that the weight of the apple caused the pointer on the scales to move, for example, we do not mean that a property of weight in general, construed as a universal, caused the moving of the pointer. Rather it was this particular weight of this particular apple that caused the pointer to move.[98]

It is to be discussed elsewhere whether he is right in insisting that not universals but particular instances of properties cause something, but let us suppose he is right in this point.[99] Let us also suppose that property instances are tropes, as the difference between property instances and tropes doesn't matter in the following discussion. Now, does Mumford's token identity theory work properly, against the argument from multiple realizability?

7-3-2 Bradley Rives' Attack on Token Identity Theory

In his paper, 'Why Dispositions Are (Still) Distinct From Their Bases and Causally Impotent', Bradley Rives attacks Mumford's token-token identity theory. He insists that dispositional property instances and categorical property instances differ in their *modal* properties and should therefore be considered distinct. He offers two examples where an object with the same disposition instance in a possible world and the actual world has a different categorical property instance from the actual categorical property instance of it in that possible world.

Let us see the first example. He takes a fragile vase V with molecular bonding α, assuming that 'the token instance of having molecular bonding α consists in the instantiation by some V's constituent atoms, a_1-a_n, of a host of highly specific properties and bonding relations, P_1-P_n, and in the resulting molecules' standing in certain bonding

---

[98] S. Mumford, *Dispositions*, pp.161-2.

[99] For a defense of such a view, see, for example, Robb (1997) and Ehring (1997, 2003). I will discuss Robb's view in Chapter 8.

relations to another'.[100] He, then, invites us to consider the following situation:

> Now, consider a possible world W in which a_1-a_n are replaced with exactly similar atoms, which instantiate P_1-P_n, and in which the resulting molecules stand in the same bonding relations as the molecules in the actual world. Let us suppose, moreover, that the replacement is carried out in such a way that there is no time at which the vase ceases to be fragile.[101]

Is the instance of *being fragile* in the possible world, W, identical with the instance of *being fragile* in the actual world? Rives insists that Mumford ought to say 'Yes'. This is because it is plausible, as Mumford himself insists it (and Rives agrees with it), that 'any two tokens with all the same causal roles are identical, and therefore are just one token'[102]. What about, then, the instance of *having molecular bonding α*? As Rives insists it, the instance of *having molecular bonding α* in W is not identical with the instance of *having molecular bonding α* in the actual world. He says:

> [S]ince in W, some of V's atoms, a_1-a_n, have been replaced with exactly similar ones, the instance of *having molecular bonding α* in W and the actual world are distinct. The reason for this is that, whatever else Mumford's property-instances turn out to be, it seems that they must be particularized entities, whose existence is tied to the very particulars that have them.[103]

As the instance of *having molecular bonding α* is possessed by the group of the particular atoms including a_1-a_n as its members, and a_1-a_n are replaced with exactly similar, but distinct, atoms, the instance of *having molecular bonding α* in W and the actual world have different possessors in each worlds, therefore they are different property instances. On the basis of the above scenario, Rives insists that Mumford ought to accept that the actual-world-instance of *being fragile* and the

---

[100] B. Rives, 'Why Dispositions Are (Still) Distinct From Their Bases and Causally Impotent', *American Philosophical Quarterly* 42 (2005), p.22.
[101] ibid.
[102] Mumford, *Dispositions*, p.162.
[103] Rives, 2005, 'Why Dispositions Are (Still) Distinct From Their Bases and Causally Impotent', *American Philosophical Quarterly* 42, p.22.

actual-world-instance of *having molecular bonding α* are distinct.

Rives argument based on the first example consists of three parts. The first part of the argument, the part which leads to the conclusion that the actual-world-instance of *being fragile* (instantiated by V in the actual world) is identical with the W-instance of *being fragile* (instantiated by V in W), is formulated as follows:

(1) The vase V would survive the replacement of a_1-a_n with exactly the same atoms.

(2) *Being fragile* is instantiated by the particular vase V.

(3) Any two disposition tokens with all the same causal roles are identical, and therefore are really just one token.

(4) Therefore, from (1)-(3), it follows that the actual-world-instance of *being fragile* is numerically identical with the W-instance of *being fragile*.

Note that the possible world W in the argument is the world where the same natural laws as ours obtain; otherwise, the W-instance of *being fragile* would have different causal roles from those of the actual-world instance, and (3) would not be satisfied. The second part of the argument, which leads to the conclusion that the actual-world-instance of *having molecular bonding α* is distinct from the W-instance of *having molecular bonding α*, is formulated as follows:

(5) The aggregate of the atoms which make up the vase V would *not* survive the replacement of its parts, a_1-a_n, with exactly the same atoms.

(6) *Having molecular bonding α* is instantiated by the aggregate of the atoms which makes up the vase V.

(7) It is impossible for a categorical property instance to be possessed by distinct particulars (substances).

(8) Therefore, from (5)-(7), it follows that the actual-world-instance of *having molecular bonding α* is distinct from the W-instance of *having molecular bonding α*.

The final part of the argument is formulated as follows:

(9) If two property instances differ in their modal properties, we can not say that they really are numerically identical. There are two distinct property instances.

(10) From (4), (8), and (9), it is concluded that the actual-world-instance of *being fragile* and the actual-world-instance of *having molecular bonding α* are distinct.

This is Rives' argument based on the first example.

One might react to the argument as follows. Mumford would accept the multiple realizability of dispositions, but this does not cause the problem for Mumford because he can say that different categorical property-instances could be identical, *in each particular object*, with disposition instances which all belong to one dispositional property. What's the difference between this case and Rives' example? The difference is this: while, in multiple realizability case, different categorical property-instances are identified with different disposition instances belonging to one disposition type, in Rives' example on the other hand, different categorical property-instances are identified with *numerically the same* disposition instance (therefore, it poses a serious problem).

We should observe that Rives relies on two criteria of identity concerning property instances or tropes, (3) and (7). (3) is the criterion of identity concerning dispositional property instances and (7) is the criterion of identity concerning categorical property instances. Let us, first, examine the latter one. (7) can be seen as stating a necessary condition for the identity of categorical property instances; for differently described properties to be identical, it is necessary that they are possessed by one and the same substance. Philosophers, who individuate tropes by means of space-time location[104], would probably reject (7), as they think that a trope keeps its identity if only it exists at the same space-time location: whether a trope is possessed by

[104] See, for instance, Schaffer, 2001, 'The Individuation of Tropes', *Australasian Journal of Philosophy 79.*

the same object or different objects doesn't matter. Anyway, the condition is quite compatible with the following plausible identity condition of tropes:

> Identity Condition of Tropes:
> Trope *a* (of type A) is (numerically) identical with trope *b* (of type B) only if (i) type A is identical with type B and (ii) trope *a* and trope *b* are possessed by the same object at the same time.

Therefore, I do accept (7) as one of the necessary conditions for trope identity.

As to another identity criterion, (3), I need to give a comment. The criterion (3) is drawn from Mumford's formulation. Mumford (1998) offers a criterion of identity between property instances. Where *d* is a variable ranging over disposition tokens, *c* is a variable ranging over categorical base tokens and *x* and *y* are variables ranging over actual and possible events, Mumford's identity criterion is formulated as follows[105]:

> $\forall d \forall c$ (($d = c$) $\leftrightarrow$ $\exists x$ (*d* causes or is caused by *x* & *c* causes or is caused by *x*)
> & $\neg \exists y$ ((*d* causes or is caused by *y* & $\neg$ (*c* causes or is caused by *y*))
> $\vee$ ($\neg$ (*d* causes or is caused by *y*) & *c* causes or is caused by *y*)))

The possible events relevant in this formulation are events in possible worlds where the same laws as the actual world obtain. Drawing from Mumford's formulation, Rives regards (3) as a criterion of identity for dispositional property instances. (3) can be interpreted as a *necessary and sufficient* condition (or at least as a *sufficient* condition) for dispositional property instances. It is, however, not at all plausible to state that if dispositional property instances (or disposition tropes) share all their causal roles then they are really numerically identical. A red trope in the apple at the left hand side of my table is not identical with another red trope in another apple at the right hand side of my table even if they share all their causal roles (suppose that they are exactly similar with each other). We, therefore, can modify (3) as follows:

(3') For any two disposition instances to be identical, it is necessary that they share all

[105] Mumford, *Dispositions*, p.162.

their causal roles.

As I see it, individuation by means of causal roles should be considered as one of the *necessary* conditions, pace Rives. Replacing (3) with (3'), Rives argument based on the first example might seem plausible. Anyway, if we take identity condition of tropes stated above for both dispositional tropes and categorical tropes, Rives' argument obtains, for the crucial point of the argument is that the disposition property instance is possessed by the same object (the same vase V) whereas the categorical property instances are possessed by different object (different aggregates of atoms).

One crucial point of the argument is that dispositional property instances and categorical property instances are never possessed by the same object. The fragility of a vase V is possessed by the vase V, whereas the molecular bonding α of the vase V is not possessed by the vase V itself, but by the aggregate of molecules which makes up the vase V. Although this is crucial for the Rives' argument, it is not at all clear why one and the same object cannot have both a dispositional property instance and a categorical property instance at the same time. We can at least say that Rives need some more argument to support this. It is also crucial that, in this example, the vase V survive its identity even if the molecules composing V are replaced with numerically different (but exactly similar) molecules. This thesis also needs some more arguments to support. Let us sum up the crucial point of Rives' argument and conceivable objections to it. In Rives' example, a dispositional trope keeps its identity through the replacement of the molecules (, for its possessor keeps its identity and the causal roles associated with the trope is also unchanged), whereas a categorical trope which is purported to be identical to the dispositional trope does not keep its identity (, for its possessor does not keep its identity). The possible objection denies either that a dispositional trope keeps its identity or that a categorical trope keeps its identity. To take the former way, we can insist that the possessor of a dispositional trope, *being fragile*, is not the vase V so that it does not keep its identity. To take the latter way, we can insist that the possessor of a categorical trope, *having molecular bonding* α, is not the aggregate of molecules but the vase V so that it keeps its identity.

That Rives is aware of *some* of these objections is shown in the following passage.

> …perhaps categorical bases are not instantiated by some of the constituents of the particulars that instantiate dispositions, but rather are instantiated by the very same particulars that instantiate dispositions. If this is right, V instantiates both *being fragile* and *having molecular bonding* α, which would preclude us from claiming, on the basis of the above scenario, that categorical and dispositional property-instances differ in their modal properties and hence are distinct.[106]

He then offers the second example to support the Distinctness Thesis. Rives invites us to consider a slightly different counterfactual scenario:

> … [C]onsider a possible world $W_\ni$ in which some of V's atoms, a_1-a_n, are rearranged so that the resulting molecules instantiate bonding relations that are slightly different from the ones they instantiate in the actual world, but ones that nevertheless realize the property *being fragile*. In the actual world, V instantiates *having molecular bonding* α, whereas in $W_\ni$, it instantiates, say, *having molecular bonding* β, a slightly different categorical realizer of *being fragile*.[107]

In this example, as Rives insists it, a dispositional trope, *being fragile*, keeps its identity, whereas two categorical tropes, *having molecular bonding* α and *having molecular bonding* β cannot in any plausible sense be considered to be identical.

But is it so apparent that the fragility of the vase V with molecular bonding α is identical with the fragility of the vase V with molecular bonding β, even if we admit, for the sake of argument, that the vase V keeps its identity through the rather radical alteration? As I see it, the vase with molecular bonding α and the vase with molecular bonding β behave differently when we probe them in detail; otherwise how can we know that one has molecular bonding α and the other has molecular bonding β? They have different causal roles, therefore, their dispositions are different as well. We may say both are fragile, in a rough meaning. But exactly speaking, their dispositions are different.

---

[106] Rives, 2005, 'Why Dispositions Are (Still) Distinct From Their Bases and Causally Impotent', *American Philosophical Quarterly* 42, p.23.
[107] Ibid.

The problem lies in the *disjunctiveness* of dispositions. It is true that Mumford avoids the overdetermination problem by identifying disposition instances with categorical property instances. But what about dispositions as type? The instance of fragility of vase A is identical with the instance of molecular bonding α of vase A. The instance of fragility of glass B is identical with the instance of crystalline β of glass B. Suppose that molecular bonding α and crystalline β are determinate and 'natural' property (in David Lewis's sense). Then the instances of α make up a class whose members are exactly similar with each other, so do the instances of β. The instances of fragility, however, cannot make up such a class, because the class that the instances of fragility make up includes, as its members, the instances of α and the instances of β with many other instances, and *ex hypothesi* an instance of α and an instance of β are not exactly similar with each other. This is a consequence from the disjunctiveness of higher-order properties. If, as PPJ takes it, dispositions are considered to be second-order properties, the disjunctiveness of dispositions is un-avoidable.

Now I would like to mention the most fundamental defect of Rives' argument (*the defect common to Humean Categoricalists' arguments in general including PPJ's*): the argument is *not* the argument *against dispositions in particular*. As I see it, the first example really concerns *constitution relation* and the second example really concerns *determinable-determinate relation*. As to the first example, it can be reformulated as follows:

> While dispositions are instantiated by *coincident objects*, their bases are instantiated not by the objects themselves but by the objects' *constituents*.
>
> If tropes are numerically identical with each other, they are instantiated by the same object.
>
> Therefore, disposition-instances cannot be numerically identical with the instances of their bases.

Therefore, as to the first example, the argument is equally applied to *any* properties which are possessed by coincident objects. So the argument is applied to a property of *being triangular* possessed by ordinary macroscopic table (*being triangular* is considered to be a typical categorical property)! Rives' second argument is really just a

modal version of PPJ's argument from multiple realizability, and it is really a problem concerning *determinable-deteminate relation* in general. The standard definition of determinables and determinates is as follows:

> An individual satisfies a determinable predicate only if it satisfies some of other determinate predicate.
>
> That an individual satisfies a determinate predicate entails (but is not entailed by) determinable predicate.[108]

Notice that Rives' second argument applies not just to dispositions but to determinables in general, such as *being coloured* (a determinate of which is *being red, being yellow*, and so on) or *being polygon* (a determinate of which is *being triangle*, *being quadrangle*, and so on). Again you can notice that *being polygon* is usually considered to be a typical categorical property.

Rives' first example points out a problem concerning constitution relation. The second example points out a problem concerning determinable-determinate relation. But neither of them has to do with the nature of dispositions. The dispositionalists are, therefore, happy to dispel the worry about the argument from multiple realizability and concentrate on more crucial problems about the nature of dispositions.

Finally, I would also like to point out that Rives' example has some plausibility only when we think about the properties of *compound* objects, such as vases, tables, and so on. If we consider the ultimate particles, particles without any constituent parts, and its properties, we are not troubled by the multiple realizability argument like Rives'. And if we can say that ultimate dispositions (and only ultimate dispositions) are causally potent, I am satisfied with it.

7-4 The Possibility of Bare Dispositions

There have been some debates concerning the possibility of 'bare dispositions'. Are there bare dispositions? To put it briefly, my answer is as follows. If 'bare dispositions' means 'dispositions that lack any basis', then the answer is no. There are no such

[108] The definition is due to W. E. Johnson. I draw this from Gillett and Rives (2005).

entities, as far as we keep the Principle of Sufficient Reason. But there is, as I see it, another meaning of 'bare dispositions'. In this sense, 'bare disposition' is a disposition whose basis is identical with the disposition itself. I think that in this sense there *are* 'bare dispositions'.[109]

To put quite intuitively, MRB requires just that when we ascribe a disposition to an object, the object must have *some property* such that the object with that property would keep the disposition in question even if it were carried to the other part of the actual world or to the other possible worlds (as long as the natural laws in those possible worlds are the same as the actual ones). It does not specify the nature of the base property itself. This abstemious requirement, however, is sufficient for the world to be understandable. In order to avoid the situations where every phenomenon occurs totally at random and we cannot predict the future based on the past and current phenomena, it is sufficient for the world to satisfy MRB.[110]

7 - 5 Against Humean Theory (The Explanation of Causal Laws)

Until this section, I have not defended the Causal Trope Theory from a main rival, Humean Theory. In this section and the next, I will argue against Humean Theory of properties. The argument has two parts. The first argument concerns the explanation of causal laws, which is the topic of this section. The second argument concerns the problem of quiddity, which I will discuss in the next section.

The first problem of Humean Theory is that it cannot explain causal laws properly. This is a traditional problem going back when Hume pointed it out. In Humean world, categorical properties are exemplified throughout the space-time: the exemplified categorical properties are scattered throughout the space-time. The problem is this: in what way we gather them up, we cannot explain the causal connection between them.

---

[109] Many philosophers seem to think that a basis of a disposition is *distinct* from the disposition itself. But I don't think this thought is well grounded.

[110] Molnar argues that fundamental physical properties attributed to elementary particles, such as charm or charge, are 'ungrounded' or 'missing base' because these particles have no parts or structures. However, bases should not be conflated with component (usually microphysical) properties. Cf. Molnar (2003), pp.131-137.

The causal theory, on the other hand, can explain the causal connection quite easily. Why does an electron with a negative charge repel another electron with a negative charge? That is because the property of having a negative charge has, essentially, a causal power of repelling the objects with a negative charge and of attracting the objects with a positive charge. How the property behaves is built in the property.

An advocate of Humean Theory might object that this solution by Causal Theory begs the question. The advocate might say that the solution just *presupposes* the causal connection, not explain it. Well, if the solution posits the causal connection *ad hoc*, then this objection would be right. As I see it, however, the causal theory does not introduce the causal connection *ad hoc* but changes the ontological framework drastically so that the problem of causal connection does not happen in the first place. This should be allowed, and this should be considered to be a good solution to philosophical problems. It might be helpful to consider another philosophical theory of causation, causal process theory, advocated by Wesley Salmon and others. Salmon introduces the causal process to tie the events that are causally related with each other.[111] If he just introduces the causal process *ad hoc*, that is, just to tie the events, then his solution to the problem of causation does not have enough plausibility. However, Salmon changes the ontological framework where the causal processes are more basic entities and particular events are *constructed* as the intersections of causal processes. He starts with causal process and then constructs other ontological items from causal processes. That is why Salmon's theory has some plausibility. The situation for the causal theory of properties is similar to that of Salmon's theory. The causal theory regards causal properties as basic. It, then, tries to explain causal laws by means of causal properties. According to the Causal Theory, causal properties are basic and causal laws are constructed out of causal properties. So we can say that its ability to explain causal laws properly should be considered as a big advantage of the Causal Theory and, conversely, its inability to explain them properly should be considered as a big disadvantage of the Humean Categorical theory.

---

[111] See, for example, Wesley Salmon, *Causality and Explanation* (New York: Oxford University Press, 1998) and Phil Dowe, *Physical Causation* (New York: Cambridge University Press, 2000) for causal process theory.

7 - 6 Against Humean Theory (The Problem of Quiddity)

The most serious problem for Humean Theory, as I see it, is the problem of quiddity. The 'this-ness' of properties is called 'quiddity'. Humean Theory has to admit the quiddity of properties. Humean Theory insists that one and the same property could have different causal powers in different possible worlds in accord with natural laws in those worlds. The property *Pn* of *having a negative charge* will repel another negative charge in the actual world, while it might attract another negative charge in a possible world, W_1, where the natural law about charge is different from that in the actual world. But how can we identify the property in W_1 with *Pn*, even if *Pn* behaves radically differently in W_1. Can we really identify a property independently of its causal powers? If we can, how?[112]

Can a defender of Humean Theory provide plausible arguments for introducing quiddity? I will examine two possible arguments for introducing quiddity and will deny both of them in the following. Before examining them, let us see two possible theories about substances. There are two theories about substances – the substratum theory and the bundle theory. The bundle theory has it that a substance is a bundle of properties (universals or tropes). The substratum theory has it, on the other hand, that a substance has, as its component, a *bearer* of those properties. The bearer of properties is called substratum or a bare particular. Now, it is instructive to consider Humean Theory of properties as the substratum theory *concerning properties* and Strong Causal Theory of properties as the bundle theory *concerning properties*. We might consider Humean Theory as introducing (so to speak) *bare properties*, in the same way as the substratum theory introduces bare particulars.[113]

Now we can admit some plausibility for the substratum theory. However, does Humean Theory have the same plausibility as the substratum theory? I will show, in the next paragraph, that the answer is negative. What I would like to show is *not* that we should take 'layer cake view' (as David Armstrong calls it) of substances and admit the bare particulars or substratum, *but* that we cannot use the same arguments for

---

[112] See R. Black, 2000, 'Against quidditism'. *Australasian Journal of Philosophy* 78, 87-104.

[113] This point is due to Chakravartty (2005).

introducing the substratum, for the purpose of supporting the plausibility of quiddity.

The first argument for introducing the substratum. There is at least one benefit for introducing *substantive* substratum. When we analyze substances by means of bundles of properties, it is very difficult to explain changes. We want to say, for example, that an apple changes its colour from green to red. However, if a green apple is a bundle of properties such as being green, being spherical, being sour, and so on, and a red apple is another bundle of properties such as being red, being spherical, being sweet, and so on, then those bundles are different things (as bundles of properties) and we cannot say that the same apple changes its colour. If we introduce an entity which sums up those properties, a bare particular, we may be able to understand the change of the colour of the same apple. Now, can we use the same argument for quiddity? Certainly not. Because we do not have to think about the situation where one and the same property changes its causal powers.

The second argument for introducing substratum. For the tropists such as C. B. Martin and Heil, tropes are the particularized ways objects are. For these philosophers, properties are not independent entities, and therefore, are essentially non-substantive even if they are particularized. Those dependent entities cannot constitute substances even if they are gathered in one bundle. We can construct a substance only when we introduce substratum as bearers of properties.[114] Can we use the same argument for introducing quiddity? Certainly not. Because we have no need to substantiate properties at all. A defender of quiddity must invent other arguments to support their view, but I cannot come up with such arguments.

There is one other worry about Humean Theory. Humean Theory comes to claim that the scientific theory *never* reach the reality of the world. The quiddity that Humean Theory posits is always beyond the reach of the scientific theory. It means not that we cannot, as a matter of fact, capture the reality, but that we can, in principle, never capture the reality of the world. Is such a theory appropriate for the theory of properties? I cannot but think that such a theory is a theory about something other than properties.[115] I will consider the problem of quiddity again with the connection of the problem of qualia in the final chapter. But let us now see how the problem of mental

[114] See Martin, 1980.

[115] Notice that the argument above applies not only to Humean Theory but also to Weak Causal Theory, as Weak Causal Theory admits quiddity of properties.

causation, the causal exclusion problem, could be solved with the general theory of properties. This is the topic of the following two chapters.

8 David Robb's Trope Identity Theory

In this chapter, I will deal with David Robb's trope identity theory. I will point out two problems concerning Robb's theory. (1) In Robb's theory, multiply realized mental types are not *genuine properties*, but just *concepts*. (2) Although 'property of causation' might be tropes as Robb insists, it should have some *connection* with types.

8-1 The Causal Exclusion Problem Reformulated

Let us make clear the upshot of the causal exclusion problem before setting out the examination of Robb's solution. The upshot of the causal exclusion problem, roughly put, is as follows. On one hand, mental properties are thought to be non-physical properties. On the other hand, the totality of physical properties seems to form the closed system. How, then, can mental properties as non-physical properties intervene in this physically closed system?: mental properties seem to be *excluded* as redundant.

David Robb characterizes the exclusion problem as a problem which arises when we try to satisfy the following three requirements at the same time[116]:

> *Relevance*: Mental properties are (sometimes) causally relevant to physical events.
> *Distinctness*: Mental properties are not physical properties.
> *Closure*: Every physical event has in its causal history only physical events and physical properties.

As to *Relevance*, we can interpret it as the following thesis which we have assumed in the preceding chapters: mental properties are causally efficacious. As to *Distinctness*, we have good reason to accept it, because we assumed, at least in some sense, that mental properties can be multiply realized. We can, therefore, easily identify *Relevance* and *Distinctness* with the premises of non-reductive physicalism in the preceding chapters. The problem is *Closure*. Discussions in 3.3 show that this principle is too

[116] Robb, 1997, 'The properties of mental causation'.

strong for a premise of an argument. It is virtually identical with Marcus's 'Closure' in section 3.3, and it obviously prohibits non-physical causes to enter into the physical domain. However, we can convert *Closure* to the weaker version of it, the Principle of Physical Causal Closure:

> [The Principle of Physical Causal Closure]: For all physical objects, if it has a cause at *t*, it has a sufficient physical cause at *t*.

We can safely use it here instead of Robb's *Closure*.

8-2 The Trope Identity Theory

David Robb proposed an ingenious solution to the exclusion problem. As Robb sees it, the three requirements appear to be incompatible because of the ambiguity of 'property' referred to in the requirements. Properties are thought of as tropes as well as types. If we distinguish tropes from types, we can satisfy three requirements at the same time.

Let us remember what was questioned about Anomalous Monism (hereafter AM). AM tried to secure the causal efficacy of mental events by identifying them with physical events. It was, however, legitimately questioned whether a *mental property* of a cause event or a *physical property* of it should be causally efficacious. Robb goes further than AM; he identifies mental tropes with physical tropes in addition to the identification of mental events with physical events. By this identification of mental tropes with physical tropes, he can, so he insists, bestow causal efficacy upon mental tropes (*Relevance* satisfied). *Closure* is also satisfied if we read 'property' in *Closure* as 'tropes'. As to multiple realizability, we can think that mental types are multiply realized by physical types (*Distinctness* satisfied). Distinguishing types and tropes, Robb interprets 'properties' in *Relevance* and *Closure* as tropes and 'properties' in *Distinctness* as types. By this interpretation, he insists he can satisfy all of the requirements at the same time.

The first impression might be that Robb's solution does not really solve the problem but just postpone it. Even if a mental trope is identified with a physical trope as Robb insists, a worry might happen again: does the trope cause the effect *as the mental*?

To this worry, Robb responds as follows. In AM, events are particulars which possess various properties (or which can be described in various ways). On the other hand, tropes are particulars which are properties in themselves (not the possessors of them). While we can legitimately ask, as to events, 'Does it cause the effect as the mental?,' we cannot ask the same question as to tropes. As Robb insists it, it is merely 'a category mistake' to ask 'In virtue of what does a trope cause the effect?'

I think that this reply by Robb has some persuasiveness. The following argument might support Robb's point. In AM, causal relata are events as particulars. If concrete particulars can be analyzed into bare particulars and attributes (cf. Loux 2002, ch.3), we would be able to think of mental events in AM as bare particulars possessing both mental attributes and physical attributes, where we would be able to ask, legitimately, which attributes are causally efficacious. In Robb's trope monism, on the other hand, mental tropes are identical with physical tropes. Can we analyze these tropes (particularized properties) into bare particulars and attributes? It should be impossible. For properties cannot lack attributes. It is mere contradiction that properties lack attributes.

However, it might be argued, against Robb, that a trope (which is both mental and physical) might have *aspects*, which could raise the same question again: which aspect of the trope is causally efficacious? This is what Noordhof raised in his 1998 paper.[117] Robb's reply to this question is that we cannot imagine aspects of properties (or properties of properties). To see which view is more persuasive, we need to examine the example Noordhof presents for his view. Noordhof says we have two perfectly clear senses in which properties can have aspects which raise questions of causal relevance.

> My house burns down. It is quite legitimate to ask which aspect of air was responsible. The answer is that the air was causally relevant in virtue of being part oxygen. So it seems that complex properties do have aspects concerning which one can ask 'Was that responsible?', namely, their constituents.[118]

Here, Noordhof seems to claim that the property of oxygen is an aspect of the property

[117] See Noordhof, 1998, 'Do tropes resolve the problem of mental causation?'.
[118] Noordhof, 1998, p.223.

of air. Or he might claim that the property of being oxygen is an aspect of the property of being air. In any case, however, the property of oxygen (being oxygen) is not an aspect of the property of air (being air); the property of oxygen (being oxygen) is the property of the constituents of air. The second example Noordhof presents is as follows;

> It also seems that, if two properties stand as determinate to determinable, we can ask 'Which is relevant?'[119]

But the relation between determinate and determinable is not the relation between properties and aspects of them. We must conclude that Noordhof does not offer a persuasive ground for his view.

8-3 Two Objections

Has the exclusion problem been successfully solved by Robb and there remains no problem? I don't think so. I will point out two problems against Robb's theory. To see first point, we ask this question. Is it possible that all the following statements obtain at the same time?

(1) A mental trope *m* is identical with a physical trope *p*.
(2) *m* belongs to a mental type *M*, and *p* belongs to a physical type *P*.
(3) Type *M* is not identical with type *P*.
(4) The members of type *M* are exactly similar with each other, and the members of *P* are exactly similar with each other.

The answer should be that it is not possible. Because exact similarity is symmetrical, transitive, and reflexive, the sets whose members are exact similar with each other are exclusive[120]. One trope cannot belong to more than two exact-similarity-types (i.e. types whose members are exactly similar with each other). Consequently, in this case, at least one of *M* and *P* should be a set whose members are non-exactly similar with each other

[119] Ibid.
[120] Cf. Armstrong, 1989, Chapter 6.

(non-exact-similarity-type).

Which of these, *M* and *P*, is non-exact-similarity-type? It should be *M*. When a mental type *M* is multiply realized by physical types *P*, *Q*, *R*, … , at least a part of the latter are considered to be exact-similarity-types. Elementary particles' mass, spin, charm etc. correspond to the exact-similarity-types. Therefore, *M* is non-exact-similarity-type.

The argument above might not be needed. When *M* is multiply realized by *P*, *Q*, *R*, and so on, *M* includes *P*, *Q*, *R*,… as subsets, and *P*, *Q*, *R* are different types with each other. Therefore, the members of *M* cannot be exactly similar with each other.

That *M* is non-exact-similarity-type means that *M* is not *genuine* property in a sense its physical realizers are. From here, it might be concluded that multiply realized mental types are just 'concepts', not properties. If Robb wants to keep both trope monism and type dualism, considering types as corresponding to *genuine* properties, Robb's solution cannot achieve it[121].

There is one way out from this, if we allow the trope has aspects. If trope *m* (= trope *p*) has aspects X_1 and Y_1, trope *m*'(= trope *q*) has aspects X_2 and Y_2, Y_1 being exactly similar with Y_2, X_1 being not exactly similar with X_2, then these two tropes are exactly similar according to aspect Y_1 and Y_2 (probably mental aspect), but not exactly similar according to aspect X_1 and X_2 (probably physical aspect). But Robb cannot take this option because Robb does not accept a sort of things such as aspects of properties.

It might be instructive to think about another possible way out. What if a cause trope is composed of more than two tropes? Suppose there are two things here – black triangle A and white triangle B. A is exactly similar with B in shape (being triangular), while A is not similar with B in colour. In the same way, if a cause trope is composed of a mental trope and a physical trope, it is possible that one and the same cause trope belongs to two distinct types each of which is exact-similarity-type. Robb cannot take this option, because it is not trope 'identity' theory any more. I think this point reveals why Robb's identity theory cannot secure mental types as genuine properties.

Let us now see the second point about the Robb's solution. Robb says that the property of causation is tropes, not types. But Robb faces the problem that the trope approach makes too many properties causally efficacious. Suppose that there is one

[121] I owe Gibb (2004) about this point.

volume trope in Ella's voice as a result of which it is true that she sings at 70dB or more, 80dB and under 90dB. Any note over 70dB will shatter the glass. Robb suggests that our inclination to say that

> Ella's voice caused the glass to shatter in virtue of being over 70dB.

is true, while

> Ella's voice caused the glass to shatter in virtue of being under 90dB.

is false should be explained by their different pragmatic implications. Robb says:

> I do not think the best way to explain this is to make types the causally relevant properties. Types simply are not the sorts of things that can be causally relevant to effects, physical or otherwise.[122]

I don't object this. But, as I see it, although it is true that types are not the sorts of things that can be causally relevant, the above case clearly shows that the property of causation should have *some connection* with types. It need not be identity. But mere pragmatic consideration is not enough. My dissatisfaction with Robb's trope monism can be boiled down to his separating classifying entities with properties of causation. We normally think that classifying entities are also properties of causation: if a ball's being 5kg makes a dent in a cushion, then being 5kg is causally relevant and all things being 5kg are classified in one group (regardless of whether the group corresponds to a universal or just a class). Robb denies this. Robb claims that classifying entities are types and properties of causation are tropes, *Distinctness* obtaining for the former and *Relevance* for the latter. This separation of classifying entities from properties of causation is difficult for me to swallow. When the two things are classified in one type, they should be similar with each other in some respect, and the similarity, it seems to me, should be connected with some kind of causal laws.

Let's take a stock. Although Robb's solution is ingenious, it has a consequence

[122] Robb, 1997, p.192

that is not very pleasant to non-reductive physicalism. According to Robb's view, mental properties (as types) turn out to be mere concepts, not genuine properties. In this sense, Robb's trope identity theory comes very close to Kim's reductionism. Also, although it might be true that types are not the properties of causation, the properties of causation should have, at least, some connection with mental types. It is not very clear how Robb's trope identity theory gives us some understanding about this matter. However, we can say this as well: although Robb's trope identity theory comes very close to Kim's reductionism, it gives us a more detailed *ontological* picture. Robb's view can be understood as giving us an ontological ground for Kim's view.

With the multiple realizability of mental properties by physical properties, it seems that we cannot have genuine mental properties, the mental properties which are qualified as real properties in the same sense that physical properties are qualified as real properties. But is it really impossible? There is a prospective theory which might allow us to have genuine mental properties. We will see it in the next chapter.

9 A Solution to the Causal Exclusion Problem: Sydney Shoemaker's View and Its Development

Sydney Shoemaker has recently offered a very attractive solution to the causal exclusion problem. In this chapter, I will examine and develop his solution. I will first describe Shoemaker's theory of properties and his view on realization. I will then examine Shoemaker's view on multiple realization and compare it with Heil's view. Although they share many insights, they differ in their views on the status of multiply realized properties. While Heil denies the existence of multiply realized properties and admits just predicates in multiple realization case, Shoemaker admits genuine multiply realized properties. I will argue that Shoemaker's view is more persuasive. After that, I will examine an objection to his view. Finally, I will develop his view.

9-1 Causal Theory of Properties

Shoemaker's solution to the causal exclusion problem is based on his own causal theory of properties. Although I have already mentioned his theory of properties in Chapter 7, it is convenient to recapitulate his theory of properties before we discuss his solution to the causal exclusion problem.

As Shoemaker sees it, properties can be viewed as sets of what he calls 'conditional powers'. His own expression is as follows:

> Any property whose instantiation can be a cause or partial cause of something will be such that its instantiation bestows on its subject a set of what I call 'conditional powers' (Shoemaker 2001, p.77)

He characterizes 'power *simpliciter'* and 'conditional power' as follows:

> A thing's having a power *simpliciter* is a matter of its being such that its being in certain circumstances, for example, its being related in certain ways to other things of

certain sorts, causes (or contributes to causing) certain effects. A thing has a *conditional* power if it is such that if it had certain properties it would have a certain power simpliciter, where those properties are not themselves sufficient to bestow that power simpliciter.[123]

For example, the property of being knife-shaped bestows on its possessor a conditional power of being able to cut wood if it is made of steel, and a conditional power of being able to cut butter if it is made of wood, and so on. Thus, the property of being knife-shaped can be considered as (or can be considered as corresponding to) the set of these conditional powers.

9-2 Formulation of the Realization Relation

Shoemaker defines the realization relation on the basis of his causal theory of properties. Properties are sets of conditional causal powers, or correspond to them. There are cases where a set of conditional causal powers is a subset of another set of conditional causal powers. A typical example is when two properties are in the relation of determinable-determinate with each other. Consider the property of being red and the property of being scarlet. In this case, the former is a determinable of the latter, and the latter is a determinate of the former. The set of the conditional causal powers corresponding to the former is a proper subset of the set of the conditional causal powers corresponding to the latter. Consider, for example, a pigeon that is conditioned to peck scarlet things but not other shades of red (e.g. pink, wine red and so on). The property of *being scarlet* has a conditional causal power of bringing about the pigeon's pecking behaviour (under suitable circumstances), whereas the property of *being red* does not have such conditional causal power as this. So the set of conditional causal powers corresponding to the property of *being red* is a subset of the set of conditional causal powers corresponding to the property of *being scarlet*.

Shoemaker defines the realization relation as follows:

[Shoemaker Realization 1] A Property *X* realizes a property *Y* just in case the

[123] Shoemaker, 2001, p.77.

conditional powers bestowed by *Y* are a subset of the conditional powers bestowed by *X*.[124]

Suppose there are four properties, *A*, *B*, *C* and *D*. Suppose also that A corresponds to a set of conditional powers {a, b, c, d}, *B* to {e, b, c, f}, *C* to {g, b, c, h}, and *D* to {b, c}. In this case, *D* is multiply realized by *A*, *B*, or *C*. Consider a brain state P_1 when a human being is in a mental state M of *having pain*. We can suppose that the pain state M could be realized by different physical setups P_2, P_3, P_4, and so on, in different organisms or robots. What Shoemaker insists is that every pain realizing physical state (P_1, P_2, P_3, …) has a common set of conditional powers (M) as its subset. For example, a conditional power of making a possessor of pain to wince when it suffers from tissue damage would belong to all the set corresponding to the physical property that realizes pain.

Defining the realization relation in this way, we get an answer to the causal exclusion problem. That is because, in mental causation, the mental property of the cause and the physical property of the cause are in *part-whole relation*: a part and the whole do not compete with each other in bringing about the effect.[125] A mental property is not excluded by the physical property which realizes it, because, in Shoemaker's framework, the former is a part of the latter and it is quite plausible to think that a part is not excluded by the whole.

Let me add some comments on it. First, we should notice that Shoemaker explicitly commit himself to an *essentialist* view on properties. According to the causal theory of properties, properties are individuated by their causal features. Therefore, if a causal law obtains, it obtains as a matter of necessity.

Secondly, according to Shoemaker's view, both the realized properties and the realizing properties could be intrinsic properties. An orthodox view on properties, advocated by Armstrong, has it that an intrinsic property can behave differently in another possible world where causal laws are different from the actual ones. According

[124] Shoemaker, 2001, p.78. The reason why this is labelled as 'Shoemaker Realization 1' is to be clarified shortly.

[125] Clapp (2001) also offers a very similar view. While Shoemaker just says that properties are *individualized by* clusters (sets) of conditional powers, Clap insists a more radical view that properties are *nothing but* clusters of conditional powers.

to the orthodox view, then, functional properties are not intrinsic properties. However, if we take an essentialist theory of properties, behaviours of a property are, so to speak, built in the property itself. Therefore, with an essentialist theory of properties, we can regard both realizing properties and realized properties as intrinsic at least in this sense (that is, they are not extrinsic with respect to natural laws or possible worlds).

Thirdly, Shoemaker's view has it that multiply realized properties are not second-order (or higher-order) properties. As we have seen when we discussed Kim's view, many philosophers advocates the view that the realized properties are higher-order properties. But as Heil plausibly argued it, this brings about the multi layered view on the world, and the reality of the entities in higher layers is always threatened (see Section 6-1). Shoemaker takes, to use Heil's word, 'a flat view' on properties and property realization.

Forthly, this being related to the third point, in Shoemaker's view, multiply realized mental properties are *genuine* properties. We don't have to consider them as just concepts or predicates as is in the cases of Kim's, Heil's, and Robb's view. We can give a full-fledged reality to the realized mental properties. A realized mental property is a genuine property in that it is a part of the genuine physical property that realizes it. In connection with this, I will discuss Shoemaker on multiple realization in the next section.

9-3 Shoemaker on Multiple Realization

What does Shoemaker say about the multiple realizability? With Shoemaker's view, we can think in the following way. Two conditional causal powers are either exactly similar with each other or not similar at all. Suppose physical properties, *A* and *B*, realize a mental property *C*. Two particular objects having *A* are exactly similar with each other. A particular object having *A* and another particular object having *B* are not as similar as the two particular objects having *A*, but they are still similar in some respect. Their not-exact-similarity is explained by the fact that they share the conditional causal powers which are exactly similar, by the fact that those particular objects share some (but less) conditional causal powers. I think this gives us a better solution to the problem. While Robb and Heil see, in a purported case of multiple realization, a set of

imperfectly similar properties in virtue of which a single predicate (e.g. 'is in pain') is applied to various objects, Shoemaker tries to find common elements among realizing properties which ground multiple realization. Robb, Heil, and Shoemaker take flat view – no higher-level properties –, thereby avoiding the threat of overdetermination or causal exclusion. Shoemaker, however, secures genuine realized properties, which is a great advantage over Robb and Heil.

We can also express their difference in this way. The difference between Heil and Shoemaker can be boiled down to their views on properties. When Shoemaker regards the property *A* as the set of conditional powers {a, b, c, d}, and the property *B* as the set of conditional power {e, b, c, f}, we can say that Shoemaker takes a *fine-grained* view on properties, although conditional powers are not themselves properties. The similarity between *A* and *B* is explained by the exact (perfect) similarity between two instances of b and another two instances of c. (Here we take particularists' view on properties, which makes the comparison with Heil's view much easier.) Heil's view, by contrast, is a *coarse-grained* one, as he insists that *A* and *B* are imperfectly similar and this imperfect similarity is a brute fact.

Which way should we take? I think that it should be, in the end, an empirical matter. But I will briefly comment on some worries with Heil's view. My worry is that similarity relation might lose objectivity in Heil's view. Heil says, 'An atomic-powered egg-beater and an apprentice chef armed with a wire whisk could be said to be functionally similar in so far as we focus on the operations of these two systems at a high level of abstraction'[126]. We could set a 'level of abstraction' high or low according to our own concern, therefore, everything resembles everything in some sense. Therefore, in Heil's view, we cannot distinguish, in principle, a case of my pain and my wife's pain on one hand with a case of my pain and an octopus pain on the other hand, or with a case of properties which are in 'family' resemblance, or even with a case of this pencil and that shirt. Shoemaker's view, by contrast, works much better here. Properties which are in 'family resemblance' never multiply realize a single property, because there is no single power common to all of them.

Moreover, scientific investigation seems to support Shoemaker's view. Suppose ultimate physics say that there exist only, say, 5 properties (a, b, c, d, e) in the world.

[126] *FOPV*, p.161.

This should mean that an instance of *b* here and another instance of *b* over there are exactly similar. It is quite unlikely that the ultimate physics poses ultimate properties such as *a*, *a'*, *b*, and *b'*, and it says that *a* and *a'* are imperfectly similar with each other. Shoemaker could take these ultimate properties as conditional powers. Some combination of these basic powers might correspond to a property, some might not. Anyway, it seems easier to make similarity relation objective for Shoemaker than for Heil. In Shoemaker's view, objects which are of 'family resemblance' with each other never realize a single property or kind; the multiple realization simply does not obtain in this case.

I, with Shoemaker, think that imperfect similarity between purported multiple realizers must be grounded by perfect similarity between causal powers which are constituents of the properties. I am happy to accept exact similarities as brute facts. If ultimate physics pose some (a finite number of) basic properties, then they are the only source of exact similarities.

9-4 Forward-looking and Backward-looking Causal Powers

Shoemaker's causal theory of properties and his view on realization has been changed slightly. The change concerns the difference between forward-looking causal powers and backward-looking causal powers. In the postscript of his 1980 paper, he refers to an counter-example which Richard Boyd pointed out to him. Boyd's example is as follows. Suppose that there are just four substances, A, B, C, and D, in the world. Suppose also that X is composed of A and B, and Y is composed of C and D. As Boyd points out, it is, at least metaphysically, possible that X and Y have exactly the same causal powers. They bring about exactly the same outputs with any possible inputs. Even if we can never distinguish X and Y by their causal features, they should be different with each other because one is composed of A and B, and the other is composed of C and D. This counter-example shows, as Boyd sees it, that there is something wrong with the causal theory of properties.

In response to this counter-example, Shoemaker introduces a distinction between 'forward-looking' causal powers and 'backward-looking' causal powers. If X and Y bring about exactly the same outcome in any situation, that means they share all

the forward-looking causal powers. However, these are not all the causal powers that X and Y possess. They also have backward-looking causal powers: X has a backward-looking causal power of being made of A and B, and Y has a backward-looking causal power of being made of C and D. X and Y are different because their backward-looking causal powers are different.

With this alteration on the causal theory of properties, Shoemaker revises his formulation of realization as follows:

> [Shoemaker Realization 2] A property *X* realizes a property *Y* just in case:
> (1) the forward-looking conditional powers bestowed by *Y* are a subset of the forward-looking conditional powers bestowed by *X*,
> and
> (2) the backward-looking conditional powers bestowed by *X* are a subset of the backward-looking conditional powers bestowed by *Y*.[127]

Notice that the part-whole relation of the sets of conditional powers is reverse in the case of backward-looking conditional powers. More specific and more determinate properties can do more things in the future; however, on the contrary, they can be produced by less things in the past. A scarlet object can cause a conditioned pigeon to peck, which a red object cannot. On the other hand, the red object can be produced more easily than the scarlet object: we can suppose that the scarlet paint is more difficult to make than the other shades of colour.

Is Shoemaker right in introducing the distinction between forward-looking and backward-looking conditional powers and in revising his formulation of realization? Concerning this point, Brian McLaughlin has recently presented an objection to Shoemaker's view. The next section I will discuss McLaughlin's objection along with the principle of proportionality.

9 - 5 The Principle of Proportionality and McLaughlin's Objection

Before discussing McLaughlin's objection, let us see another aspect of Shoemaker's

[127] See Shoemaker (2007).

view, the principle of proportionality. The principle of proportionality is the principle that causes must be *proportionate* to effects.[128] Let me explain it by means of Yablo's vivid example[129]. Suppose that a pigeon is conditioned to peck red things (not just scarlet things but also pink things and wine red things and so on) and that she, in fact, is pecking a scarlet thing. What is doing the causal work (or what has causal efficacy) in this case is, as Yablo sees it, not scarletness but redness of the object, as the object would have been pecked by the pigeon even if it had been pink instead of scarlet. Causes must be *proportionate* to effects, according to Yablo. The property of being scarlet is 'too large' for the cause of the pecking effect. Yablo formulates the principle of proportionality as follows:

> [The Principle of Proportionality] *c* causes *e* only if (i) *c* is not screened off by any of its parts, and (ii) whatever has *c* as a part is screened off by it.[130]

Note that the part-whole relation mentioned here is *intensive* (not extensive) part-whole relation: a typical example is the relation between determinates and determinables. Note also that a cause *x* 'screens off' *y* from an effect *e* if and only if *e* would have occurred even if *x* had occurred without *y*.[131]

It is important to notice that Shoemaker's solution to the causal exclusion problem, examined in the preceding section, just guarantees that a physical property and a mental property does *not conflict* (or *not compete*) with each other. It just guarantees that a whole does not pre-empt a part of itself. It does not, therefore, positively state that mental properties *do indeed* cause something in some cases. However, if we require the principle of proportionality and if it is possible to show that mental properties are more proportionate to the effects than physical properties according to Shoemaker's theory, then we can say that metal properties do cause the effects in appropriate cases according to Shoemaker's theory. And Shoemaker indeed seems to be able to take this way, for the set of the (forward-looking) causal powers bestowed by a mental property is a part

---

[128] The principle of proportionality is forcefully defended by Yablo in his discussion of mental causation. See Yablo (1993).
[129] See Yablo (1993).
[130] See Yablo (2007), 'The Seven Habits of Highly Effective Thinkers,' http://mit.edu/~yablo/effthink.html.
[131] Ibid.

of the set of the (forward-looking) causal powers bestowed by the physical property and this means that the mental property is an *intensive* part of the physical property that realizes it.[132]

But there might be a stumbling block. In a recent paper, Brian McLaughlin attacks Shoemaker's formulation of realization.[133] As McLaughlin sees it, it is crucial for Shoemaker that physical properties *entail* mental properties: the fact that an object has a physical property P_1 entails the fact that the object has a mental property M if P_1 realizes M. Without entailment of mental properties by physical properties, Shoemaker encounters two difficulties: (1) If entailment fails, supervenience also fails, which means that he must give up physicalism; (2) If entailment fails, he loses the only resource to guarantee the principle of proportionality.[134] However, as McLaughlin sees it, his formulation ([Shoemaker Realization 2]) fails to do it. That is because the part-whole relation in the second part (concerning *backward-looking* conditional causal powers) of [Shoemaker Realization 2] is such that the set of conditional causal powers bestowed by the realizing property is a part of the set of conditional causal powers bestowed by a realized property. Suppose that P realizes M. [Shoemaker Realization 2] has it that the set of the forward-looking causal powers of M is a subset of the set of the forward-looking causal powers of P. Therefore, considering only the forward-looking causal powers, the fact that an object has P seems to entail the fact that the object has M. However, as the second part of [Shoemaker Realization 2] states that the part-whole relation between P and M is reverse as to backward-looking causal powers, considering both the forward-looking causal powers and the backward-looking causal powers, the fact that an object has P does *not* entail the fact that the object has M. Therefore, according to [Shoemaker Realization 2], realizing properties do not entail realized properties.[135]

What should we make of this? As I see it, if we count on backward-looking conditional causal powers for indentifying properties, then McLaughlin's objection works. I am not so sure, however, if backward-looking conditional causal powers have

---

[132] Cf. Shoemaker (2001), p.78.
[133] McLaughlin (2007), 'Mental Causation and Shoemaker-Realization,' *Erkenntnis* 67.
[134] See McLaughlin (2007). We should also remember that according to Johnson's classic definition, the determinables-determinates relation is defined by means of logical entailment relation. See section 7-3-2.
[135] Ibid.

something to do with identification (or individuation) of properties. If two properties, *P* and *Q*, share all the forward-looking conditional causal powers, but differ only in their backward-looking conditional causal powers, is it not the case where *P* and *Q* are, in fact, the same property (or exactly similar with each other) but they are produced in different processes? I am quite doubtful that Boyd's counter-example really describes the real possibility. If *X* and *Y* share all the forward-looking conditional causal powers, then should we not conclude that *X* and *Y* are identical even if one is made of *A* and *B* and the other of *C* and *D*? It might be the case where a material is made of different constituents due to a mysterious chemical reaction.

I cannot present a conclusive argument as to if the backward-looking causal powers play a crucial role in identifying properties. I can only say that McLaughlin's argument poses a very interesting point about the identification of properties and that if we need the principle of proportionality, we should reject [Shoemaker Realization 2] and go back to the first one, [Shoemaker Realization 1]. As I see it, it is quite doubtful that backward-looking conditional causal powers have something to do with the identification of properties. I believe, anyway, that there are many interesting points to be considered around here in the future research.

Finally, I would like to just briefly mention if Shoemaker should take *reductive* causal theory, the *strongest* causal theory according to the classification I made in section 6-3-1. As we saw in 6-3-1, it seems that Shoemaker once took reductive causal theory but later altered his view. As I see it, Shoemaker should take reductive one for the following reason. If the set of causal powers bestowed by *M* is a part of the set bestowed by *P* and properties *are exhausted* by their causal powers as the reductive causal theory insists, then it is natural to regard *M* as an intensive part of *P*, and therefore we seem to be able to avoid the conflict or competence between *P* and *M* (because a part and the whole do not compete with each other and the whole does not pre-empt a part of itself). But if properties are not exhausted by their causal powers and we should add something other than their causal powers to the properties, then is it so natural that *M* is an intensive part of *P*? In that case, we cannot so easily state that the former is a part of the latter. Rather we might have to say that the former, *M*, is *distinct* from the latter, *P*. For even if the sets of causal powers are in part-whole relation, the

added something to *M* and the added something to *P* might be totally different entities with each other. For this reason, I think Shoemaker should take *reductive* causal theory. The argument is not conclusive at all, but it should be clear that here also there are many interesting points to be developed in the future research.

All in all, I think that Shoemaker's view, with necessary amendments, is the most persuasive approach so far. We might have to give up [Shoemaker Realization 2]. We also have to consider if Shoemaker's view should be *reductive* causal theory or not. But the advantage of his view is clear: it gives us a clue to insist that mental properties are *genuine* properties and yet have their causal powers distinguished from their physical realizers. And throughout these considerations, I believe I have shown that the metaphysical study of properties has a great and direct impact on the mental causation debates. Now let us depart from the problem of mental causation and get into the problem of consciousness.

１０ Physicalism and Consciousness

The conscious experience of human being is usually considered to be the most serious problem for physicalism. Why is consciousness a problem for physicalism? In order to understand this, we should see how it is that we understand a phenomenon in a physicalistic world view.

Let us see a formulation of physicalism again (for details, see Chapter 2). According to Frank Jackson's formulation, physicalism insists that any world which is a physical duplicate of our world is a duplicate *simpliciter* of our world.[136] To use the concept of supervenience, it insists that every property in our world supervenes on micro physical properties and relations, the properties of fundamental particles and the relations between them. This is the second requirement for physicalism discussed in 2-3.

With this rough formulation of physicalism (the second requirement for physicalism), let us ask what it is to understand a phenomenon (a phenomenon in general) in a physicalistic way. Let us take up a physicalistic understanding of biological gene, for it is comparatively uncontroversial that we can properly understand a biological concept of gene in a physicalistic way. First of all, how do we grasp the concept of gene? We begin with noticing that parents and their children are alike in some ways, in other words, that some traits of organisms are conveyed from parents to their children; this is the genetic phenomenon. Examining genetic phenomena in more detail, we come to think that if we *posit* a kind of *particle* which serves as a unit in genetic phenomena, we can systematically understand the genetic phenomena; we call this particle gene. Now, when we try to understand this concept, gene, *physicalistically*, what must be done? We must show that (1) genes are nothing over and above physical particles and (2) these physical particles play a causal role which is the constitutive character of gene. We know that DNA plays a causal role of gene by specifying how proteins are to be composed. We know that what genes do and we know that DNA do what genes do. In this way, we understand genetic phenomena in a physicalistic way.

[136] F. Jackson, *From Metaphysics to Ethics*, p.12.

This is usually called functional explanation.

Functional explanation, in general, takes the following form. First, a phenomenon which is to be explained is shown to be or have a functional property (or state) – whatever has a causal role essentially. Next, it is shown that a physical entity plays, as a matter of fact, the causal role. When a phenomenon is functionally explained in this way, we get a physicalistic understanding of the phenomenon. Notice, here, the connection between functional explanation and the formulation of physicalism offered above. If we can give a functional explanation to every fact in our world, then we can say that any world which is a physical duplicate of our world is a duplicate *simpliciter* of our world; every property in our world supervenes on micro physical properties and relations. Furthermore, we can say that the functional explanation explains why this formulation obtains. If we cannot give a functional explanation of a fact, it is difficult to believe that a physical duplicate of our world is a duplicate *simpliciter*, or that every property in our world supervenes on micro physical properties and relation, unless we have another good explanation.

Can we, then, understand consciousness physicalistically? What matters here is the qualia of conscious experience. There are two types of argument which try to show that we cannot understand conscious experiences with qualia within a physicalistic world view. One type of the argument relies on the conceivability of certain situations. Another type concerns the knowledge of conscious experiences. Let us take up the former first.

10-1 The Conceivability Argument and the Representation Theory of Consciousness

The Conceivability Argument

Some philosophers who believe that we cannot solve the hard problem of consciousness within a physicalistic world view, rely on the argument from conceivability. Consider two people, *A* and *B*, in exactly the same functional state. We seem to be able to imagine that *A* has a quale which normally occurs when we see blue objects, while *B* has another quale which normally occurs when we see yellow objects. In this case, both

A and *B* would reply 'It's blue' when they see the sky and are asked its colour, for we suppose that their functional states are identical. This possible situation is called 'inverted qualia' situation. We also seem to be able to imagine that *A* has qualia and *B* lacks any qualia. This is called 'absent qualia' situation. There is a more extreme case, where *A* and *B* are in exactly the same *physical* state (not merely the same *functional* state) although *A* has qualia and *B* lacks any quale; this is called 'Zombies' situation.[137] Some philosophers argue that we can not get a physicalistic understanding of qualia because we *can conceive* of these cases and the cases seem to point out there are always some aspects which evade physicalistic description of the world. This is the outline of the conceivability argument against physicalism.

For the argument to be plausible enough, the argument must presuppose that the conceivability entails metaphysical possibility. We can conceive some situations, while we cannot conceive other situations: we can, for instance, conceive that the sky is red, while we cannot conceive that the sky is blue *and* red. It seems plausible that this conceivability corresponds with the metaphysical possibility, the possibility of the way the world could be. The sky could be red, but the sky could not be blue and red at the same time. Considering whether some situations are conceivable or not is always determined under some descriptions of the world, we can state this principle as follows:

> (CEM) Conceivability of a situation in *some* description of the world entails that the situation is metaphysical possible.

Let us call the principle, CEM (Conceivability Entails Metaphysical possibility). CEM seems plausible enough. If we can say some situations are metaphysically possible and other situations impossible, how can we say that other than considering some situations as conceivable and other situations as inconceivable?

If we accept CEM, then it follows that we cannot accommodate qualia in a physicalistic world view. For if CEM is right, then the conceivability of Zombie entails the metaphysical possibility of Zombies, which means that it is quite possible that one person has qualia and the other thing (Zombie) with exactly the same physical state as that person lacks any quale. As this denies the second requirement for physicalism, we

[137] See D. J. Chalmers, *The Conscious Mind*, Chapter 7.

cannot retain a minimal physicalism. So goes the argument.

A **Reply from A Posteriori Necessity**

To reply to the conceivability argument, the physicalists should deny CEM in some ways. A notable reply from the physicalists concerns a case of *a posteriori* necessity (let us call it 'A Reply from A Posteriori Necessity'. It goes like this. At first glance, it seems that we are able to conceive that water is not H_2O but XYZ, as the conceivability argument insists. We seem to be able to conceive that the liquid which is called 'water', which is colourless and transparent, and which freezes at 0 degree centigrade (and so on), is not H_2O but, in fact, XYZ. At least, it seems possible that a people who, lacking proper scientific knowledge, does not know that water is H_2O, can conceive that water is not H_2O but XYZ. However, the reply says, according to our current science, water *is, by definition,* a particular chemical material with a particular molecular structure H_2O, and it is, in fact, *not* metaphysically possible that water is not H_2O but XYZ even though we seem to conceive of the possibility of it. The upshot is that some necessity is *a posteriori*: although we come to know, by empirical investigation, that water has a microphysical structure, H_2O, this does not mean that the fact is mere contingent. There are cases where a posteriori necessity holds, and the fact that water is H_2O is one of those cases. This reply, thus, concludes that conceivability does not entail metaphysical possibility.

Does this reply successfully establish a case where conceivability does not entail metaphysical possibility? As I see it, this reply is on the right track. But one might object to this as follows. One might cast doubt on this reply, using the distinction which David Chalmers draws between primary intention and secondary intention.[138] One might argue against the reply by insisting that the purported counterexample of CEM relies on the ambiguity of the concept (or word), 'water'. As Chalmers sees it, we must distinguish two meanings (intentions) of 'water'. 'Water' might refer to an entity with colourlessness, transparency, liquidity, and so on (whatever properties we normally associate with water in our world); 'water', on the other hand, might also refer to an entity with a particular molecular bonding which realizes such properties as colourlessness, transparency, liquidity and so on in our world. In the former sense

[138] Cf. Chalmers, *The Conscious Mind*, pp.131-140.

(Chalmers calls it 'primary intention'), 'water' could have any micro structure so long as it has colourlessness, transparency, liquidity and so on in the world which it belongs to. But in the latter sense (Chalmers calls it 'secondary intention'), 'water' refers to the material which has those properties in our world. We can say that if we distinguish primary intentions and secondary intentions, the conceivability and the metaphysical necessity goes together. Understanding 'water' in the former sense (primary intention), we can conceive that what has colourlessness, transparency, etc. is not H_2O but XYZ; there is a metaphysically possible world where XYZ realizes such properties as colourlessness, transparency, etc. according to the natural laws obtaining in that world. Understanding 'water' in the latter sense (secondary intention), on the other hand, we cannot conceive that a molecular structure which realizes water properties in our world (that is, H_2O) is not H_2O, and we can plausibly insist that this situation is not metaphysically possible either. Therefore, the reply which recourses to *a posteriori* necessity is not conclusive.[139]

I doubt, however, that Chalmers' argument in the previous paragraph is the right way to support the conceivability argument and to rebut the physicalism. I agree with him that if we distinguish primary intentions and secondary intentions, the conceivability and the metaphysical necessity goes together. However, I don't see how exactly this supports the conceivability of inverted qualia situations, absent qualia situations, or Zombie situations. To support the conceivability argument against physicalism, it should be possible that we conceive an inverted qualia situation (and others) with the intention of physical properties and qualia restricted to either primary or secondary respectively. I don't see how it is possible.

Anyway, it should be admitted, at least, that the reply from a posteriori necessity is not conclusive. And the argument so far suggests that we will have to revise the concept of consciousness in some way if we want to accommodate consciousness within a physicalist world view. I will expound this approach in the next section.

Revising the Concept of Consciousness

It is suggestive to consider the concept of life. Former days, people believed that we

[139] Cf. ibid.

cannot understand the concept of life physicalistically. They believed that although the bodies of organisms are composed of physical materials, such as atoms or molecules, the physical materials themselves could never be alive; some non-physical energy must be added to the physical materials for them to be alive. Nowadays, however, many people believe that we can understand the concept of life physicalistically. Many people and philosophers believe that we do not need some mysterious energy anymore. What caused this change? We can point out two factors. First, the concept of life became clearer. We now know, thanks to the advancement of the biological science, that the essence of life consists in, for example, some functions such as self-replication, intake of energy, metabolism, and so on. Second, we now know, thanks to the advancement of the biological science again, that the physical materials, such as atoms and molecules, can realize these essential functions – that is to say, the physical materials occupy the causal roles specified by the functions essential to life activities. We should particularly notice here the first factor. Just as we come to believe that we can understand life activities physicalistically with the change in the concept of life itself, so we might be able to expect that we can understand consciousness physicalistically with the appropriate revision in the concept of consciousness itself. We need to clarify the conscious phenomena and to establish the proper concept of consciousness.

If we describe the world by appropriate concepts, then what is conceivable by our concepts would be metaphysically possible. However, the concepts that we currently possess to describe the world would not be fully appropriate; at least we should not expect that they are. If some of our concepts are not appropriate, then what is conceivable might turn out to be metaphysically impossible under those concepts. What we must do in these cases is to revise our old concepts so that we can make explicit the discrepancy between the conceivability and the metaphysical possibility. The concepts we now possess are not appropriate for the proper description of the world in these cases, so we must revise the concepts. The case of life is an example that we revised the concept in such a way. It is quite plausible that, in the case of consciousness, we face the same problem as we faced when we tried to understand the concept of life physicalistically. It might be true that we can conceive Zombies in our current concept of consciousness. But it is not quite certain that our current concept of consciousness is fully developed and appropriate to describe the world. In fact, there are some evidences

that apparently cast doubts on our current concept of consciousness.

One example is the blindsight. There are the patients who have blindsight. They suffered from tissue damages in some parts of their brains by accident. They deny that they see a visual image (spots on the screen, for example) which would be recognized by them if the damaged part of the brains worked properly. They cannot recognize that there are spots on a part of the screen. However, when they are asked to guess the place of the spots or the direction to which the spots move, they can correctly answer the question without being self-aware of perception. To be precise, they can not always but *sometimes* answer the tests correctly, the possibility of their answering correctly being higher than when someone who doesn't see the screen at all makes a guess. This is the case of blindsight; they can, in some way, perceive something without knowing themselves that they perceive those things. It is sometimes said that the case of blindsight suggests the possibility of Zombies. Judging solely by their function, we cannot distinguish the patients who have blindsight from the people who have normal visual ability and sometimes make mistakes in the test of spot detection. Judging by their consciousness, however, they are different. While the people with normal visual ability have consciousness, a quale of the colour of spots or the shape of spots, when they detect spots on the screen, the patients with blindsight do not have such consciousness.[140] Dennett offers a thought experiment in the blindsight case. He asks what if a patient with blindsight receives substantive trainings and gets an ability to answer the test correctly without exception. The patient would not have such feeling as making mere guesses 'blindly'. Should we really regard this patient as not having any consciousness when they answer the test perfectly?[141]

These cases suggest that our current concept of consciousness is not appropriate, is confusing, and needs some revisions. The conceivability argument has it that our current conception of consciousness allows us to conceive Zombies and our current physical theory cannot capture this aspect of the world. Replying this argument, we can say that if we revise our current concept of consciousness and perhaps our current physical theory as well, we would be able to capture all aspects of the world including conscious experiences.

---

[140] For many cases of the blind sight, see, for example, Ramachandran (1999).
[141] See Dennett (1991), Chapter 11, Section 2.

The Representation Theory of Consciousness

In the preceding section, we saw the conceivability argument against physicalistic understanding of consciousness. We argued that the reply from a posteriori necessity is on the right track though it is still open to further discussions. We also saw that even if Zombies are conceivable under our current concept of consciousness, this does not immediately deny the possibility of physicalistic understanding of consciousness, for it is very likely that our current concept of consciousness is inadequate and needs some revisions. Our next task is to offer a positive account of *how* we can understand consciousness physicalistically. Even if physicalism is true as an ontological doctrine, we currently do not have with us an understanding of how physicalism is true. In the following, I will survey what I believe the most prospective approach to physicalistic understanding of consciousness – the representation theory of consciousness, which is powerfully advocated by Gilbert Harman and other philosophers.

We often regard qualia as non-physical properties. But, as Harman sees it, it is not right way to see qualia. Consider our experience of seeing a red apple. We are liable to think that what has a red quale is this experience. But, insists Harman, reflecting on this experience, what has a red quale is not this experience but the apple itself outside us. Harman generalizes this insight by using the concept of intentionality. Our experiences have intentionality: our experiences are representations. As to a representation, we must distinguish the properties of the representation itself, i.e. the intrinsic properties of the representation on one hand and the properties that the represented objects (the intentional objects) have on the other. The sentence printed on a paper, 'The apple on the table is red', has many intrinsic properties, such as being black, being composed of seven words, etc. The intentional object, the apple on the table, also has many properties, such as being red, being round, etc. The intrinsic properties of representations and the properties of intentional objects are usually different. Once this difference being noticed, as Harman sees it, the quale that we experience should be regarded as the property of the intentional object, not the property of our experience; qualia are to be understood as the properties of intentional objects.[142]

[142] See G. Harman, 'The Intrinsic Quality of Experience', in J. Tomberlin, ed.,

Harman's insight is important for physicalistic understanding of qualia. First, it suggests the way to revise our current concept of consciousness and qualia that brings about problems posed by the conceivability argument. Second, it suggests, more specifically, a way to understand qualia physicalistically via a physicalistic understanding of intentionality. We already have some physicalistic theory of intentionality and it is usually considered that a physicalistic understanding of intentionality would be easier than a physicalistic understanding of qualia (I will discuss physicalistic theories of intentionality later)[143]. If we can understand the concept of intentionality physicalistically, and the concept of consciousness can be captured by its intentional character, then we will have a way to understand consciousness physicalistically. This project is called representation theory of consciousness.

Harman's insight on qualia and the project of the representation theory of consciousness are still rough sketches: it just outlines a way to physicalistic understanding of consciousness and there are several problems to be solved. Firstly, Harman just analyses *some* conscious experiences by means of the concept of intentionality and shows that the qualia accompanied by those experiences are the properties of the intentional objects represented by those experiences. If we want to get a physicslist theory of consciousness, we should have to show that *every* conscious experience can be analyzed in the same way. Secondly, and obviously, not every representation is a conscious experience. The sentences on this page are representations, but they are not conscious experiences themselves, although they induce conscious experiences on readers. The physicalists have to draw a line between representations which are connected with conscious experiences and mere representations which are not connected to conscious experiences, and also have to specify the criterion to draw the line. Let us see these problems.

Are Every Conscious Experiences Analysed Intentionally?

Let us consider if Harman's view could be extended to every conscious experience. An

---

Philosophical Perspectives 4, California: Ridgeview.

[143] For physicalistic theories of intentionality, see, for example, the following studies: F. Dretske, *Naturalizing the Mind*; R. G. Millikan, 'Biosemantics', *Journal of Philosophy* 86, pp.281-97.; M. Tye, *Ten Problems of Consciousness*.

anti-physicalist would refute that although some qualia might be regarded as properties of intentional objects, there should be other conscious experiences which don't have any intentional content. Many philosophers agree that perceptual experiences and sensations have representational aspects. Consider, for example, the perceptual experience I undergo when I see a red apple on the table. This experience represent that a red round object is on a brown plane. It also seems uncontroversial that some feelings such as emotions and moods have representational content. There are, however, some feelings which seem to lack representations.

It is sometimes said that experiences of pain are not representational. One might insist that pain does not represent anything. Suppose one feels headache. It seems that the quale accompanied with this experience does not represent anything and that the quale is an intrinsic property of the experience itself (not a property of an intentional object such as table or an apple). This objection, however, could be replied. When we feel pain, we experience the pain as located somewhere in our body; we feel headaches in our heads. Sensations could be analysed within Harman's framework in exactly the same way as visual experiences; the only difference between them is that the former has, as its intentional objects, a subject's body.

What about the other mental states than pain? Michael Tye persuasively argues that the qualia accompanied with emotional experiences are understood as the compound states or properties of the subject's body. The qualia that we have when we feel angry can be understood as the compound states or properties of the body, the components of which are the body being in the state of excitement, muscles being stretched, and the face being flushed etc. Imagery experiences can be understood as imitational visual experiences. The qualia that we have when we consciously think can be understood as imagery experiences of sounds and characters we use when we think.[144]

Are Every Representations Conscious Experiences?

Next, let us consider the second problem. If the representational theory of consciousness is right, then conscious experiences are representations. But are all representations

[144] See Tye (1997), pp.125-131.

conscious experiences? Obviously not. The things such as characters and pictures on a paper represent something. But it is absurd to think of these as conscious experiences. What conditions must be obtained for representations to be conscious experiences?

Two types of approaches have been suggested. The first type requires that representations must have *higher order* representations – the representations of the first order representations. This approach is based on the intuition that when an experience is conscious, the subject of the experience *is aware of* its own mental state. To use a more physicalistic image, for an experience to be conscious, the subject of the experience must *monitor* what is going on in the system. For example, when I drive a car, I am mostly aware of my mental states. But sometimes, especially when I am too familiar with the road, I just get to the destination without being aware of which road I have taken or how exactly I drove to the destination. In this case, my mental states when I was not aware of them should be non-conscious. This type of approach is further divided into two groups. One suggests that the higher order representation must be perception of the first order representation. According to this theory (Higher Order Perception theory, or HOP theory for short), a mental state is a conscious state if and only if there is a perception of the first order mental state. An experience of an apple is conscious if and only if the subject of the experience perceives the first order mental state.[145] The second group also requires the higher order representations, but it requires that the higher order representation must be higher order *thought*. According to this theory (Higher Order Thought theory, or HOT theory for short), a brain state representing a red apple is a conscious experience if and only if this brain state is represented by a higher order thought with a content, 'I am seeing a red apple'.[146]

But these attempts are not free from difficulties. It is well known that there are two kinds of consciousness – phenomenal consciousness and access consciousness, the former being virtually equal to qualia and the latter being expressed as follows:

> Access Consciousness: A state is access consciousness if it is poised for direct control of thought and action. To add more detail, a representation is access

[145] For an earlier formulation of this theory, see Armstrong, *A Materialist Theory of Mind*, Chapter 5.
[146] See D. Rosenthal, 'Two concepts of consciousness,' *Philosophical Studies* 49, 1986: 329-359.

> consciousness if it is poised for free use in reasoning and for direct rational control of action and speech. An access state is one that consists in having an access representation.[147]

For the first type of attempt, we would be able to point out that phenomenal consciousness might be confused with access consciousness. As Chalmers points out, our concepts of consciousness are usually ambiguous and sometimes include the meaning which is different from phenomenal consciousness.[148] What we really need here are the conditions for a representation to be a *phenomenal consciousness*. But are the attempts above give the conditions for other consciousness than the phenomenal one that is functionally definable?

We might be able to say at least access consciousness is explained by Higher Order theories. But what about phenomenal consciousness? According to HOP theory, both first order perceptions and higher order perceptions include non-conceptual contents. On the other hand, according to HOT theory, while the first order perceptions include non-conceptual content, higher order perceptions include conceptual content. Now as to HOP theory, we can ask what is higher order non-conceptual content? Are higher order perception not collapsed into the first order perception? Also, according to HOP theory, there must be perceptual organs which perceive the first order perception. But what exactly are they like?

As to HOT theory, also there are many questions to be asked. Firstly, according to HOT theory, we have to have concepts as many as we can distinguish phenomenal qualia. But do we really have so many concepts? Secondly, according to HOT theory, higher order thought makes the first order perception conscious. But are higher order thoughts themselves really conscious? If they also need to be thought by much higher order thought, then obviously we face a regress. Thirdly, it seems that the requirement of HOT theory seems to be too strong. According to HOT theory, it is necessary for a representation to be a conscious experience that a subject has an ability of conceptual thought or has language in some way or another. This means that animals other than human beings could not have any consciousness unless they have languages or they are

---

147 The definition is from Block (1995).
148 See Chalmers (1997), Chapter 1.

capable of conceptual thought. Furthermore, we must say, according to HOT theory, that babies or even adults who lack linguistic ability or conceptual thought ability cannot have any consciousness. These are quite counter-intuitive. If they maintain their view, they have to show that what must be revised is our intuition about consciousness; this, however, seems to be very difficult.[149]

What Has the Representation Theory of Consciousness Achieved So Far?

It is true that there are many problems left for the representation theory of consciousness to solve. The suggestion in the previous section is just a sketch and we need to do more research about the representation theory of consciousness. However, it is also important to see what the theory achieved so far. If we take the representation theory, we have a prospect of soothing a fear of conceivability argument. According to the representation theory of consciousness, qualia are properties of intentional objects. The quale which Adam has when he perceives a red apple is nothing but the property of what Adam's experience represents. And what Adam's experience represents is determined by the intentional content of the representational state (a brain state of Adam's) which realizes Adam's experience. Now if we have a physicalistic account of intentionality (as we will discuss later in this chapter), the content of the representational state (a brain state of Adam's) is determined by the fact that the brain state of Adam's satisfies a causal role or a functional role. Therefore, if we take the representation theory of consciousness and a physicalistic account of intentionality, then there is no possibility that Adam has a different quale than he actually has. One cannot conceive of that possibility.

This is true whether or not we take a causal theory of properties (pursued in Chapter 6 and 7). If we take a causal theory of properties, a physical state has a causal role in all possible worlds: it is metaphysically necessary that a physical state (property) has a causal role which is essential to that state (property). Even if we do not take a causal theory of properties, we can still say that a physical state (property) has a causal role in all possible worlds where the same natural laws as ours obtain. Therefore, if we

[149] For a very clear and accessible survey of higher order theories, see Kim (2011), p.283-289.

confine ourselves in those nomological possible worlds, we can say that the same physical state (property) realizes the same causal (functional) role. In any case, if two subjects are in the same physical state, then they must have the same representational state and, thus, have the same conscious experience in all nomological possible worlds. We can conclude, therefore, that if we take the representation theory of consciousness, we have, at least, a good perspective to cope with the problems posed by the conceivability argument.

Now the next task would be to find out a proper physicalistic understanding of intentionality. But before that, we have to see another very popular argument against physicalism – the knowledge argument.

10-2 The Knowledge Argument and the Representation Theory of Consciousness

The Knowledge Argument

There are two arguments against physicalistic theory of consciousness: one is the conceivability argument and the other is the knowledge argument. As to the conceivability argument, we argued above that the argument does not completely refute the physicalist pursuit in the theory of consciousness. We also argued that if the representation theory of consciousness is on the right track, then we do not have to be bothered with the possibility of inverted qualia situation, absent qualia situation, or Zombie situation. Now we must consider the second argument, the knowledge argument.

This well known argument, persuasively put especially by Frank Jackson, focuses on the knowledge about the experiences. Frank Jackson starts with the following formulation:

Physicslism (FJ): Every knowledge is physical knowledge[150]

[150] In fact, Jackson uses the term 'physical information' rather than 'physical knowledge'. See Jackson (1982). I believe this alteration does not affect the argument here.

With this in mind, he offers a thought experiment where a scientist, Mary, try to know the visual experiences. Mary was born and brought up in the circumstance where everything is black or white. She has a proper perceptual system to perceive colours, but she has not seen coloured things in all her life. She is a scientist and has all physical and physiological knowledge concerning colour perception. The only thing she lacks about colour perception is an *experience* of seeing colours. She has all knowledge on physical and functional facts about seeing coloured things. However, if she gets out of the black and white circumstance and sees coloured things, it seems that she will get new knowledge about colour perception, that is, what it is like to see coloured things. It seems that the knowledge she gets when she sees, say, a red thing, is something she cannot get until she, in fact, sees a red thing. This seems to show, as Jackson insists it, that there is non-physical knowledge that cannot be reduced to the physical one. Jackson calls this argument the knowledge argument. The argument can be formulated as follows:[151]

(1) Physicalism (FJ): Every knowledge is physical knowledge.
(2) Before released from black-and-white room, Mary has all the knowledge about human visual system.
(3) When she is released and she sees a red thing, she gets new knowledge.
(4) Therefore, what she gets is not physical knowledge.
(5) Therefore, there is knowledge other than physical knowledge, and Physicalism (FJ) is false.

Now we have to notice, first, that Physicalism (FJ) should be a form of *epistemological* physicalism, not *metaphysical* physicalism. Compare with the following formulation:

Physicalism (Metaphysical): Every fact is physical fact.

It might seem, then, that the knowledge argument concerns only epistemological

[151] Ibid.

problem and it has nothing to do with metaphysical matters which we are interested in throughout this thesis. But it is not so obvious.

Let us ask, here, what sort of knowledge Mary lacks according to the argument. The knowledge argument presupposes that Mary can know all the physical and functional facts about perceptual experiences. This means that Mary would know all perceptual facts which is expressed by ordinary propositions; let us say that Mary has all the 'propositional knowledge'. For example, Mary knows that 'red induces excitation on the subject of the experience', because the excitation of the perceptual subject and the process of inducement of excitation can be understood as objective physical states – appropriate functions. The knowledge that Mary lacks, then, would be something like what can only be expressed, very vaguely, 'To be a bat is such and such'. Or it should be said that 'to be a bat is such and such' is not propositional knowledge at all, but it should be a 'non-propositional knowledge'. Now, some proponents of the knowledge argument might insist that the reason why Mary cannot have non-propositional knowledge, is because there exists knowledge about *non-physical facts*. Being put in this way, the knowledge argument, based on an *epistemological* formulation of physicalism – Physicalism (FJ) – might have *metaphysical* implications or consequences. We will get back to this problem, later, when we discuss 'Reply from Different Ways of Knowing'.

Also, even if we take the representation theory of consciousness and we understand all physical and functional facts about consciousness, it seems that we still cannot clearly understand how it is for Mary to see red things. At least in this respect, we need to give some explanation of the knowledge argument.

So, how can we reply to the knowledge argument? The argument is obviously valid, therefore we must ask either

(a) Are premises true?

or

(b) Even if the argument is correct, does the knowledge argument implies something about *metaphysical* physicalism?

I will take up, at first, a reply which questions (a) – The Ability Hypothesis. Then I will take up a reply which questions (b) – Reply from Different Ways of Knowing.

The Ability Hypothesis

The first reply questions premise (3) in the argument. It insists that what Mary gets in (3) is not propositional knowledge (knowledge about facts) but an *ability* to perceive red or other colours. As Ryle emphasized it, we can distinguish two kinds of knowledge: the knowledge of 'know that' and of 'know how'. Consider a man who knows every physical fact about a bicycle but cannot ride on it. Even though he knows the mechanism of a bicycle and how the energy produced by a foot is transmitted from the pedal to the wheel, he cannot ride on a bicycle. In this case he has much knowledge of 'know that' but does not have the knowledge of 'know how'. In other words, he does not have an *ability* to ride on a bicycle.

Some philosophers argue that the knowledge argument confuses these two kinds of knowledge. When Mary is released from the black and white room and sees a red thing for the first time, what she acquires is not the knowledge about the non-physical facts, but an ability to recognize the experience of seeing a red thing. This is 'know how'. Before she gets out of the room, she cannot remember the experience of seeing a red thing (as she doesn't have the experience yet). When she gets out of the room and sees a red thing for the first time, she still cannot identify *this* experience she is having with the experience of seeing a red thing until she is told that it is. After she gets out of the room, sees a red thing, and is told that what she is looking at is a red thing, she can recognize the experience and remember the experience later. She can also distinguish the experience of seeing a red thing with the experience of seeing a green thing. It is not a mystery that she does not have these abilities before she gets out of the room and acquires them when she has the experience of seeing a red thing and others. She does not acquire the knowledge of non-physical facts about the world.[152]

At this point, a proponent of the knowledge argument might respond like this. We could admit that Mary gets an ability, or 'know-how' knowledge, in (3). But the fact that one gets an ability does not exclude the possibility that one gets a propositional knowledge as well. In this case, Mary certainly gets an ability, but we can still insist that Mary gets a propositional knowledge of seeing a red thing as well. To this, a

[152] See, for example, Laurence Nemirow, 'Physicalism and the Cognitive Role of Acquaintance', in Lycan ed., *Mind and Cognition: A Reader*, 1990; David Lewis, 'What Experience Teaches', 1988.

proponent of Ability Hypothesis (i.e. physicalists) can ask: what kind of propositional knowledge does Mary gets, then? It should be something like 'a red thing appears to me *such and such* (or like *this*)'. This shouldn't seem to be an ordinary propositional knowledge as it contains indexicals, such as 'such and such' or 'this'. The proponent of the knowledge argument would might respond: that the proposition includes indexicals is exactly the character of non-physical knowledge. At this state, the debates seem to come very close to metaphysical arena.

There are other problems with Ability Hypothesis. Why is it that Mary cannot acquire the abilities without having the experiences? Why does she need some experience in order to acquire the abilities? If the reason why she needs the experience is that non-physical properties of the experience play some crucial roles, then the Ability Hypothesis turns out to be inadequate for a defence of physicalism. The proponents of the ability hypothesis must explain the need of the experience in a physicalistic framework. I will get back to this point later, when we discuss a reply from the representation theory.

Although, as I see it, Ability Hypothesis is on the right track, there are many problems to be solved. And as I see it, the representation theory of consciousness gives us a better explanation on this.

Reply from 'Different Ways of Knowing'

As indicated in the previous section, the knowledge argument might have metaphysical consequences, if the existence of non-physical knowledge (non-propositional knowledge) implies the existence of non-physical *facts*. Many advocates of the knowledge argument seem to take this way. This is because, so it seems, they presuppose that for each fact, there is just one way of knowing it. If this is true, then we easily infer that there are two kinds of facts, physical facts and non-physical facts fromt the fact that there seem to be two kinds of knowledge.

Paul Churchland, however, denies this. As he sees it, there are two ways of knowing a single fact. And for Chrchland there are only physical facts. He accepts the functionalist view on conscious experiences and takes it that the conscious experiences are nothing over and above the functional states that are realized by the brain states. He

argues that we can know the brain states in two ways. On the one hand, we can know the brain states by scientifically observing the brain. On the other hand, we can know the brain states by *introspection* as well. Suppose one sees a red thing at time *t*. That person's brain state at *t* can be known by scientifically observing that person's brain. It is also known to that person when he/she introspects his/her own mental state at *t* and recognizes that he/she is then seeing a red thing. The former way of knowing is available to Mary, but the latter way of knowing is not available to Mary. As Churchland sees it, when Mary goes out of the black and white room and gets to know what it is like to see a red thing, what she acquires is not knowledge of a non-physical fact but a way of knowing what she already knew.[153]

Is this reply to the knowledge argument persuasive? I don't think so. The main reason for this, as I see it, is that Churchland's reply is *not* compatible with the representation theory of consciousness. Churchland's reply has it that the knowledge that Mary acquires when she first has the experience of seeing a coloured object, is the knowledge about her own brain state. If the representation theory of consciousness is true, all the properties that we have when we experience something are the properties of the intentional objects. Therefore, all the knowledge that we acquire when we have some experiences should be the knowledge about the intentional objects of the experiences, not the knowledge about the brain states. The representation theory of consciousness, then, has it that Mary discovers some new facts about the world when she first has the experience of seeing coloured things, whereas Churchland's view has it that Mary just uses a new way of knowing the old facts. This incompatibility itself does not refute the replies by Churchland, but we can at least say that if one wants to reply to the knowledge argument in the way Churchland does, one needs to offer a theory of consciousness as persuasive as the representation theory of consciousness.

A Reply from the Representation Theory of Consciousness

According to the representation theory of consciousness, when one has an experience,

[153] See, for example, Paul Churchland, 'Reduction, Qualia, and the Direct Introspection of Brain States', *Journal of Philosophy* 82, 1985, p.8-28.

what she *sees* in the experience is not the experience itself but the intentional objects and their properties that the experience represents. That is to say, she does not see her inner mental states but the world outside her. Therefore, as the representation theory of consciousness has it, what Mary acquires when she is released from the black and white room is not knowledge about her own experience, but knowledge about the world that her experience represents. Mary has all the knowledge about the colour perception but she still lacks some knowledge or information *about the world* before she is released from the room. It is true that she may know that the apple on the table in the normal room next to her black and white room is red by listening to someone talking on the telephone or by reading a note describing the normal room. But if she perceives the apple in the normal room, she will know a new knowledge or information about the redness. She will acquire a kind of knowledge which she cannot acquire by linguistic or conceptual ways. This should be non-propositional knowledge or non-conceptual knowledge.

What Mary acquires is not just non-conceptual knowledge, however. Before Mary gets out of the room, she can use the word 'red' in some proper ways, for she learned the meaning of the word 'red' by reading books or having talks on it. But she cannot apply the word 'red' to her own experience. When she gets out of the room, sees a red thing directly by her eyes, and is told that what she is looking at is a red thing, she acquires the non-conceptual knowledge *and* the knowledge about the connection between the non-conceptual knowledge and the conceptual knowledge as well: she will be able to get a conceptual knowledge that the book on the floor is also red by a perceptual experience of seeing a red book on the floor.

Remember the problem that worried the ability hypothesis. The ability hypothesis does not offer an explanation of how Mary can acquire the abilities to recognize and remember the experiences. The representation theory of consciousness offers an explanation: Mary acquires these abilities by acquiring the non-conceptual knowledge and the knowledge about the connection between the conceptual knowledge and the non-conceptual knowledge.[154]

[154] Thanks to Suzuki (2004) for suggesting this.

１０-３　How to Analyze Intentionality in a Physicalistic Framework

So far, I defended the representation theory of consciousness and qualia. According to the representation theory, the fact that our conscious experiences have qualia is explained as follows: a conscious experience is a mental state possessed by a particular in a subject; the mental state is a representation; qualia are the properties of the represented objects (the intentional objects). Although the representation theory has some problems of its own to be solved, I advocate the theory as an appealing option. The main reason for my advocate is that we have a prospect of explaining intentionality within a physicalistic framework. In this section, I will deal with the problem of understanding intentionality within a physicalistic framework.

When we perceive, believe, image, expect, or remember, our minds are *directed* to something. Pictures or linguistic expressions are also *directed* to something – something that they mean. The property of being directed to something is called intentionality. At this state, we can presuppose that 'something' could be any metaphysical items: substances, properties, events, facts, and so on. When one sees an apple, one's perception is directed to the apple; when one believes that the apple is sour, one's belief is directed to the fact that the apple is sour. Also, 'something' does not have to exist or obtain; one might fear a ghost even if it does not exist; one might believe that the earth is flat even if the fact does not obtain in the actual world.

When we want to keep a physicalistic worldview and accommodate the intentionality therein, we face some difficulties. First of all, intentional objects are sometimes non-existent, or at least non-actual as stated above. How can a mere physical thing be directed to something non-existent or non-actual? Some philosophers think that the ultimate physical properties won't include intentionality as a brute property. Jerry Fodor says as follows:

> I suppose that sooner or later the physicists will complete the catalogue they've been compiling of the ultimate and irreducible properties of things. When they do, the likes of *spin*, *charm*, and *charge* will perhaps appear upon their list. But *aboutness* surely won't; intentionality simply doesn't go that deep. It's hard to see, in face of this consideration, how one can be a Realist about intentionality without also being,

> to some extent or other, a Reductionist. If the semantic and the intentional are real properties of things, it must be in virtue of their identity with (or maybe of their supervenience on?) properties that are themselves *neither* intentional *nor* semantic. (Fodor, *Psychosemantics*, p.97)

This quite well shows a motivation for physicalists to look for a physicalistic understanding of intentionality and a possible strategy such physicalists could take. If a physicalist doesn't want to deny the reality of intentionality, it would be a pressing problem for her to accommodate intentionality, in some way or other, in a physical framework.

There are, roughly speaking, three approaches to the reductive explanation of intentionality in a physicalistic framework: causal approach, functional approach, and teleological approach. In the following I will briefly review those approaches and show that teleological approach is the most prospective, although, as will be insisted at the end of this chapter, the three approaches might cope with each other within a physicalistic understanding of intentionality.

The Causal Approach

First, let us see the causal approach. This approach, advocated by the philosophers such as J. Fodor and F. Dretske, relies on causal correlations (in some sense) between representations and represented objects. Noticing an ordinary fact that a perception of, say, a duck is caused by the duck, this approach insists that mental states represents what cause those representations. Thus, it is formulated as follows:

(1) *X* represents *Y* if and only if *Y* causes *X*.

This is the crudest version of the causal approach. Being crude, it immediately faces some problems. When a mental state represents a duck, the representation (call it D-representation) does not have to be caused by a duck: a rabbit could produce a representation about a duck. Also, even if a representation is caused by a duck, it might not be a representation about a duck: the mental state might represent a rabbit.

A more elaborated version of this approach is as follows:

(2) *X* represents *Y* if and only if the occurrence of *X* is causally dependent on *Y*.

However, even if elaborated like this, the causal approach does not completely overcome the problems. First, some representations can represent something that does not have any causal relation with the representations. For example, we can think about non-existent objects (i.e. we can have a representation that is directed to non-existent objects) such as unicorns and Sherlock Holmes, and it is obvious that non-existent objects do not cause (or are caused by) physical objects: non-existent objects cannot have any causal relation with anything. We can also think about mathematical entities, and it is not clear at all how mathematical entities enter into causal relations with physical objects.

Furthermore, and more importantly, there is another very serious problem – the problem of error. (2) insists that *X* being causally dependent on *Y* is necessary and sufficient for *X* representing *Y*. Suppose, first, causal dependence is necessary for representation. Then *X* could not represent *Y* when *Y* does not exist. Therefore, it is not possible that one *mis*represents a duck. Suppose, next, that causal dependency is sufficient for representation. Then *X* would represent anything on which *X* is causally dependent. Therefore, D-representation in the first paragraph of this section would represent a disjunctive content, 'rabbit or duck'. In any case, we cannot explain the possibility of error in representation.

The Functional Approach

The functional approach applies functionalism, a popular view in the philosophy of mind, to intentionality. According to the functionalism in the philosophy of mind, a mental state of a subject is determined by a pattern of the causal connections the mental state has with sensory inputs, behavioural outputs, and other mental states of the subject. The functional approach expands this analysis to contents of mental states. It, thus, insists that a pattern of the causal connections a mental state has with other mental or physical states determines not only the mental state but also a *conten*t of the mental

state. In the case of contents, 'the causal connections' is interpreted wider: it includes not only ordinary physical connections but also inferential connections. In this way, the functional approach tries to analyse intentionality by means of functions.[155]

Sadly, though, the functional approach faces virtually the same problem as the causal approach faces – the problem of error. This is because a pattern of the causal connections of a mental state of a subject includes only inferences that the subject *in fact* does. Therefore, the pattern cannot capture inferences that the subject *should* do.

The problem of error shows that intentionality has a normative aspect. A mental state that represents a duck (D-representation) is not merely caused by a duck; D-representation is rather a state that *should* represent a duck. If D-representation is caused by something other than a duck, there occurs an error: D-representation *mis*represents a duck. The normativity cannot be captured by mere causal correlations. Even if we introduce the concept of 'ideal conditions' and try to specify the causal correlations which determine the representation relation, it is not very clear if we can explain what is ideal condition without relying on normativity.

The Teleological Approach

Can we explain the normative aspect of representation in a physicalistic framework? The teleological approach is prospective in this respect.

This approach takes notice of the biological organs. The biological organs of living organisms, such as hearts and lungs have a particular function that fits the purpose of those organisms' survival or reproduction. The long neck of a giraffe fits the purpose of the giraffe' nibbling leaves on high branches, thereby contributes to the giraffe's survival and reproduction in the circumstances where many trees are high. Lungs have a function of bringing oxygen in, thereby make it possible that our body combust nutrition; in this way, the function of our lungs fit the purpose of our survival and reproduction.

The biological organs, thus, have a function that fits the purpose of the possessors' survival and reproduction. What is more, the living organisms have been naturally selected in virtue of their having *those* functions. These functions are usually

[155] See for example, Block (1986) for this approach.

called *teleological function* or, as R. G. Millikan calls them, *proper function*.

The teleological approach for mental representations has it that mental representations have a kind of teleological functions. Biological organs have their own biological functions and how a biological organ works is explained by a biological function of the organ. In the same manner, how a mental state represents is explained by a biological function of the mental state.

A distinct advantage of the teleological approach, compared with the first two approaches, is that the teleological approach has a prospect to explain the normative aspect of intentionality: proper functions are considered to have normative character. A proper function is determined not by what the possessor of the proper function in fact does (or is disposed to do), but by what it *should* do for the possessor to survive and reproduce itself. For example, even though almost all sperms fail to be inseminated with ova, sperms' proper function is still to be inseminated with ova. Even if an animal's heart might be disabled and cannot pump blood properly, the proper function of the heart is still to pump blood.

So if we analyse mental representations and intentionality by means of biological functions, then we might be able to accommodate intentionality (especially its normative character) within a physicalistic world view, as we obviously have physicalistic explanation of biological functions. There is, however, one stumbling block. Although mental states can be considered to have some functions that contribute to the subject's survival and reproduction, those functions are not always the function *in virtue of which the organisms have been naturally selected*. For example, a belief that there is a snake causes the behaviour of escaping from the snake, and thereby contributes to the survival of the subject. But this belief itself is not passed down from generation to generation. What is passed down is not a mental state itself, but a mechanism that produces and uses this mental state. The belief-producing-mechanism produces, for example, a belief that there is a snake in case there is a snake ahead and a belief that there is an apple in case there is an apple in front of the subject, thereby contributes to the subject's survival and has been naturally selected: it contributes to the subject's survival by producing the appropriate beliefs according to the various environments. The belief-using-mechanism uses a thus produced belief to conduct practical reasoning and bring about the appropriate behaviours, thereby contributes to

the subject's survival and has been selected. In this way, belief-producing-mechanisms and belief-using-mechanisms have teleological functions. A particular belief itself, on the other hand, does not have teleological functions, as it is not naturally selected in the history of evolution. Although beliefs themselves do not have teleological functions, they can have a kind of functions that derive from genuine teleological functions. The teleological function of the whole mechanism, the combination of belief-producing-mechanism and belief-using-mechanism, is to produce a belief P and to use *P* in order to conduct practical reasoning and bring about the appropriate behaviours in case of *P*. In this case, we can say that *P* has a kind of teleological function derivatively. Millikan calls the former type of function a *direct* proper (teleological) function and the latter a *derived* proper (teleological) function.[156]

Proper Function and the Normativity of Intentionality

Let us see Millikan's definition of proper function in detail. A device or an item of a system (an organism) has a (direct) proper function as a member of a special kind of family that she calls a 'reproductively established family'. A (direct) proper function is defined as follows. Where *m* is a member of a reproductively established family *R* and *R* has the reproductively established character *C*, *m* has the function *F* as a direct proper function if and only if:

(1) Certain ancestors of *m* performed *F*.
(2) In part because there existed a direct causal connection between having the character *C* and performance of the function *F* in the case of these ancestors of *m*, *C* correlated positively with *F* over a certain set of items *S* which included these ancestors and other things not having *C*.
(3) One among the legitimate explanations that can be given of the fact that *m* exists makes reference to the fact that *C* correlated positively with *F* over *S*, either directly causing reproduction of *m* or explaining why *R* was proliferated and hence why *m* exists.[157]

[156] Millikan (1984), p.41.
[157] Millikan (1983), p.28.

An example would make it easy for us to grasp the idea. Your heart has a function of pumping blood. Your ancestors' hearts had the pumping blood function in the past, and the pumping blood function had positive effects for survival and reproduction of the ancestors in the history of evolution. This is why you have your heart now. In this case, the function of pumping blood is a proper function.

Notice that all three conditions (1)-(3) refer to the *past* facts. (1) requires that your ancestors' hearts pumped blood in their bodies. (3) requires that the existence of your heart is explained by the historical fact that the family was naturally selected in virtue of the correlation between the constitution of the human hearts and the function. Notice also that the character which plays the function (constitution of ancestors' hearts) might not be the same as the character the device in fact has (constitution of, say, your heart). Your heart might lack the character of accomplishing blood pumping and still it has the function of blood pumping: this is the case that a device with a function fails to accomplish the function. A proper function depends not on the property of the device, but on the history of the family where the device belongs. Although a heart produces sound of pulses, producing the sound of pulses has nothing to do with the history of the heart being naturally selected. Therefore it is not to be regarded as a proper function of the heart. Thus we can say that a proper function is determined not by what the device in fact does, but by what it *should* do (or what it is designed to do). In this way, Millikan explains the normativity by referring to the history.

More about Teleological Approach

Teleological functionalism insists that the normativity of representations is captured from a viewpoint of biological functions. However, as Millikan emphasizes it, the mere insistence that what a mental representation represents is determined by what the metal representation has a function to represent, is quite vague unless we have a theory of mental representation.[158] There are two possible ways to take concerning theories of mental representation within teleological functionalism. In this section, I will examine

[158] See Millikan, 1989, 'Biosemantics', reprinted in her *White Queen Psychology and Other Essays for Alice*, p.84.

them.

Let us see an example by Dretske. Dretske says:

> Some marine bacteria have internal magnets (called magnetosomes) that function like compass needles, aligning themselves (and as a result, the bacteria) parallel to the earth's magnetic field. Since these magnetic lines incline downwards (towards geomagnetic north) in the northern hemisphere (upwards in the southern hemisphere), bacteria in the northern hemisphere … propel themselves towards geomagnetic north. The survival value of magnetotaxis (as this sensory mechanism called) is not obvious, but it is reasonable to suppose that it functions so as to enable the bacteria to avoid surface water. Since these organisms are capable of living only in the absence of oxygen, movement towards geomagnetic north will take the bacteria away from oxygen-rich surface water and towards the comparatively oxygen-free sediment at the bottom.[159]

The problem is what the direction of internal magnets represents. Does the direction of internal magnets represents 'the geometric north' (the direction of earth's magnetic field)? Or does it represent 'the absence of oxygen'? Dretske takes the first option.[160] Dretske basically takes a version of the causal approach for representations. He insists that there must be causal connections, in some way or other, between representations and intentional objects. And he relies on teleological functions just for explaining the normative aspect of representations. His theory being based on the causal connections between representations and intentional objects, Dretske insists that contents of representations are determined by the causes of the representations. The cause of the movement of internal magnets of the marine bacteria is the earth's magnetic field, therefore, the direction of the internal magnets of the marine bacteria represents the geomagnetic north. Just as a length of mercurial column in a thermometer indicates the temperature of the environment, so the direction of the internal magnets indicates the geomagnetic north of the environment.

On the other hand, some advocates of the teleological approach such as Millikan,

---

[159] Dretske (1986) p.26.

[160] Other than Dretske (1986), see, for example, Neander (1995) for an approach to this direction.

Papineau, Elder, and Price, take the second option.[161] They take it that not only causes but also effects of representations concern the contents of the representations. The internal magnets are directed to the geomagnetic north, and thereby are directed to the absence of oxygen, which is crucial to their lives. Thus, the function of their internal magnets should be to move themselves towards the absence of oxygen. And the internal magnets represent absence of oxygen. If the function of the internal magnets is to indicate the geomagnetic north, then we cannot understand why they have such function.

As Millikan sees it, when a representation represents something, there must exist a *user* (or a *consumer* in Millikan's terminology) of the representation.[162] That is to say, Millikan states the following as a necessary condition for intentionality:

> [Necessary Condition for Intentionality (NCI)]: For a mechanism to be representational, it is necessary that the mechanism has a function to control the consumer of the mechanism in the way that the consumer's behaviours conform to the condition in the environment.

As to the marine bacteria, their propulsion unit (i.e. ciliation) is a consumer of the internal magnets. If the internal magnet has a function to control the propulsion unit in the way that the unit moves the bacteria to the absence of oxygen, then the internal magnet is qualified as a representational mechanism. Thinking in this way, we cay say that the content of the representation (the internal magnet) concerns the absence of oxygen, not the geomagnetic north. That is because for the behaviour of the consumer (the propulsion unit) to be optimum, it is necessary that the direction of the internal magnets corresponds to the absence of oxygen. The internal magnets contribute to the survival of the bacteria because the direction that they point to corresponds to the absence of oxygen, and thereby the internal magnets have a proper function of controlling the propulsion unit in that way.

Now let us see how Millikan's teleological approach can explain the normativity of representations. The first point to be noticed is the introduction of proper functions

---

[161] Cf. Papineau (1993), Elder (1998), and Price (2001).
[162] Millikan (1993), p.88.

and the definition of proper functions refers both to organisms which can reproduce themselves and to the history of organisms' reproduction. If there is anything normative in a physical world, the most probable candidate should be a 'success' of reproduction of organisms. I cannot even imagine what other option we have. We will get back to this later in this chapter.

The second point to be noticed is the introduction of *consumer*. As Price pointing out, the introduction of consumer mechanism gives us the source to explain the problem of error.[163] If we consider only two kinds of items, representations on one hand, and intentional objects on the other hand, then it is very difficult to say something is right or something is wrong; there is less possibility that errors occur. Adding consumers to representations and intentional objects, there come to be communication between representations and consumers; therefore, representations are considered to convey some information. If they can convey information, they can convey erroneous information as well. Thus, in Dretske's view, the bacteria that are taken out to the presence of oxygen (danger zone) by a bar-magnet are not 'deceived', whereas, in Millikan's view, they are certainly 'deceived'. This is certainly an advantage of Millikan's view.[164]

The Problem of the Indeterminacy of Content

In general, when one tries to explain the content of an intentional state by the function of that intentional state, one often faces the problem that the content is not uniquely fixed because the function is not uniquely fixed. The advocates of teleological approach insist that if we use *teleological* functions, not functions *tout court*, to explain intentionality, we will not be bothered about the indeterminacy of content, and this is the advantage of this approach compared with the other approaches. However, there has been offered some cases where even teleological functions seem not to be fixed uniquely. In this section, I will examine the objection of this kind to the teleological approach and see if the teleological approach can overcome the objection.

Consider the predatory activities of frogs. When frogs see small flying insects,

[163] Price (2001), p.77.
[164] I am helped in this paragraph by Price (2001) and Maeda (2004).

they stick out their tongues and catch small insects to eat them. However, it is not only small insects that cause frogs to stick out their tongues and catch them: it is observed that when frogs see something small and blackish (for example, a small piece of blackish plastic) moving quickly, they stick out their tongues and try to catch them. Now, given a moving small blackish object, a frog's neural mechanism forms a neural state *N*. The neural state *N*, with other necessary conditions, causes the frog to stick out its tongue. What, then, does *N* represent? Does it represent that there is a small insect flying ahead? Or does it represent that there is a moving small blackish piece in ahead? Or both?

Our intuition seems to insist the first option. The neural state *N* represents that there is a small insect flying ahead. If there is indeed a small insect flying in front of a frog, the frog's neural state *N* *rightly* represents the insect and leads the frog to successful actions (e.g. catching the insect and eating it). If there is, instead, a small black piece flying in front of the frog, *N* *wrongly* represents the piece as an insect and leads the frog to unsuccessful actions. Therefore, we would like to say that *N* represents that there is a flying insect: if *N* is indeed produced by an insect, *N* represents rightly; if *N* is produced by a small black piece, *N* represents wrongly.

Despite our intuition, however, the neural mechanism of the frogs is such that the neural state *N* can be produced when there is a small black piece flying in front of the subject; the frogs' neural mechanism is not such that *N* is produced only when there is a flying insect in front – it is not so precise. Therefore, even if *N* is produced by a small black piece, it does not mean that there is something wrong with the frog's neural mechanism. Even in this case, the frog's neural mechanism does not malfunction; it works properly. If so, doesn't *N* represent that there is a moving small black piece?

Although, intuitively, *N* seems to represent that there is a flying insect in front, the frog's *N*-producing-mechanism seems to suggest that *N* represents that there is a moving small black piece in front. Which content does *N* represent? The teleological approach insists that if we introduce teleological functions, we can fix the content of the representation in accordance with our intuitions. As the teleological approach has it, it is important to distinguish two mechanisms: a mechanism which *produces* a representation and a mechanism which *uses* ('*consumes*' to use Millikan's terminology) a representation. In frogs' case, we can think of the former as the frog's perceptual

mechanism, and the latter as the frog's motion mechanism.

On the one hand, the motion mechanism of the frog uses the representation *N* and makes the frog to stick out its tongue and try to catch the objects. The motion mechanism has contributed, in the biological history of frogs, to the survival and reproduction of the frogs, and the teleological function of the frog's motion mechanism is to make frogs to stick out their tongues and catch a flying 'insect': for it does not contribute to the survival or reproduction of frogs at all to make frogs to catch a moving piece of plastic. On the other hand, as to the frog's perceptual mechanism, its teleological function is to support the motion mechanism to conduct the teleological function of the motion mechanism. The frog's perceptual mechanism supports the frog's motion mechanism by producing a neural state *N*. In order to support the motion mechanism properly, the perceptual mechanism must produce *N when there is a flying insect* in front of the subject. If the perceptual mechanism produces *N* when there is a moving piece of plastic, it does not have the teleological function of supporting the motion mechanism to conduct its own function of making the frog to catch an insect. Therefore, if the perceptual mechanism of the frog has a teleological function at all, the teleological function must be to produce *N* when there is a flying 'insect' in front of the subject.

Now, *N* itself has a *derived* teleological function. The derived teleological function of *N* is to be produced when there is an insect flying in front and to control the frog's behaviours. Therefore, we can say that *N* represents not that there is a moving small black piece but that there is a flying insect.

In this way, the teleological approach seems to solve the problem of indeterminacy of content quite well. Jerry Fodor, however, argues that even if we introduce teleological functions, mental contents are not uniquely fixed because teleological functions themselves are not uniquely fixed. Fodor insists that we could attribute a teleological function to an organ in a different way than Millikan does. Millikan insists that the teleological function of the frog's motion mechanism is to lead frogs to catch a flying insect and eat it. But if frogs inhabit in a *normal* environment, the environment should be such that all the moving, small and black objects are, in fact, flying insects. Therefore, in the normal environment, to catch a moving small black piece should be nothing but to catch small insects. Then we can say that the frogs'

motion mechanism contributes to the frog's survival and reproduction by leading the frog to catch a moving small black piece. We can also say that the function of the frogs' perceptual mechanism is to produce *N* when there is a moving small black piece. In this way, therefore, we would be able to say that *N* represents that there is a moving small black piece in front.[165]

What exactly is at issue here? If the environment where the frogs have been developed so far in their biological history is the environment where there exist moving small black pieces which are *not* flying insects, then the teleological functions of frogs' motion mechanism and perceptual mechanism should be about flying insects (not moving small pieces in general), because the moving mechanism and the perceptual mechanism have been naturally selected not because they perceive and react moving small black pieces in general, but because they perceive and react a more specific subset of moving small black pieces, flying insects. However, what Fordor has in mind shouldn't be like this. What if the frogs have been developed in the environment where there are no moving small black pieces other than flying insects? That is to say, what if the extension of a set of moving small black pieces and a set of flying insects are exactly the same with each other. If this is the case, the teleological function of frogs' motion mechanism and perceptual mechanism would be not only about flying insects but also about moving small black pieces. If all moving small black pieces are insects, then to make frogs catch and eat moving small black pieces is nothing but to make frogs catch and eat flying insects. If to make frogs to catch and eat flying insects contribute to the frogs' survival and reproduction, then to make frogs to catch and eat moving small black pieces also contribute to the frogs' survival and reproduction. Therefore, if to make frogs catch and eat flying insects is the teleological function of the motion mechanism, then to make frogs catch and eat moving small black pieces is also the teleological function of the motion mechanism. In exactly the same way, to produce *N* when there are moving small black pieces is nothing but to produce *N* when there are flying insects. Therefore, if to produce *N* when there are moving small black pieces is the teleological function of the perceptual mechanism, then to produce *N* when there are moving small black pieces is also the teleological function of the perceptual mechanism. In sum, it seems to be concluded that even if we introduce teleological functions, *N* can

[165] Cf. Fodor (1990) pp.69-77.

be interpreted not only as representing that there are flying insects but also as representing that there are moving small black pieces: the content of *N* is not determined. Intuitively, however, the content of *N* seems to be fixed even in this case. This should be what Fodor intends to point out.

C. L. Elder tries to solve this problem by noticing the intensional character of causal explanation[166]. The argument is as follows. There is a well-known fact that the causal explanations form intensional contexts. As Donald Davidson insists it, the context '*c* causes *e*' is usually considered to be extensional. Consider a sentence expressing a causal relation, 'The storm caused the falling of the oak tree'. If 'the storm' and 'the event reported on the newspaper' refer to the same event, then substituting 'the storm' with 'the even reported on today's news paper' does not change the truth value of the sentence. The sentences expressing causal relations form extensional context. On the other hand, the sentences expressing causal explanations form intentional context. Consider a sentence expressing causal explanation, 'The storm causally explains the falling of the oak tree'. If we substitute 'the storm' with 'the event reported on the newspaper', then the truth value of the sentence changes from true to false. Now, the fact that a frog catches a flying insect causally explains the fact that the frog takes the nutrition. However, although 'a flying insect' and 'a moving small black piece' have the same extension, the fact that a moving small black piece does not causally explain the fact that the frog takes the nutrition. This is the reason why catching a flying insect is the teleological function of the motion mechanism of the frog, whereas catching a moving small black piece is not.

One might insist, however, that appealing to the intensional character of causal explanations is not appropriate. 'A flying insect' and 'a moving small black piece' have the same extension in the actual environment where frog's have been developed. And in *this actual environment* that a frog's catching a moving small black piece *does* explain the frog's taking nutrition in this environment. Confining the situations to the environment that the frogs in fact have been developed, there obtains a causal law that if a frog catches a moving small black piece, the frog takes nutrition. It is, of course, *possible* that the environment that the frogs have been developed is the environment where the moving black small pieces are not always insects, and in this possible

[166] See Elder (1998).

environment the causal law does not obtain. However, as long as we consider the teleological functions of the frog's motion mechanism and perceptual mechanism, we should exclude these possibilities. If we are concerned with the possible (counterfactual) teleological functions of the frogs, then the possibility matters. As we are concerned with the actual teleological functions of the actual frogs, what matters is the environment that the frogs have actually been developed in their actual biological history. We are now supposing that the environment where the frogs have actually been developed is such that the extension of 'moving, small, and black pieces' and the extension of 'flying insects' are identical with each other. As long as we confine the environment to this one, the law 'if a frog catches a moving small black piece, the frog takes nutrition' obtains in this situation. If this law obtains, then the frog's catching a moving small black piece causally explain the frog's taking nutrition. And we can regard making frogs to catch a moving small black piece as the teleological function of the frog's motion mechanism. Therefore, appealing to the intensional character of causal explanations does not solve the problem of indeterminacy. Appealing to the intensional character is to appeal to a counterfactual environment, but the facts in those counterfactual environments do not matter the problem about the teleological functions in the actual environment.[167] Fodor offers an example that might support this point.[168] Consider the brightly coloured fish that are found in sunless ocean deeps. The bright colour of the fish might have some functions advantageous to their survival and reproduction if their environment were lit up. However, the bright colour does not have such functions in the environment where they actually inhabit, ocean deeps. What teleological function an organ has is determined by in what environment it has actually been developed and how it has actually contributed to the owner's survival and reproduction. What function it has in counterfactual environments doesn't matter at all. One might attack Millikan and Elder in this way.

Is this objection conclusive? It is true that teleological functions of an object should be functions in the actual world, not in other possible worlds. The teleological functions of frogs' perceptual mechanism and motion mechanism should be those in the actual world where the extension of 'a flying insect' and 'a moving small black piece'

[167] Thanks to Nobuhara (1999) for pointing out this objection.
[168] See Fodor (1990).

are the same. However, in order to specify actual functions, we need to consider possible situation. Even if '*X*' and '*Y*' have exactly the same extension in the actual world, that does not immediately mean that '*A* causes *X*' and '*A* causes *Y*' have the same status as causal statement. Suppose every instance of *X* are followed by every instance of *A*, and, also, every instance of *Y* are followed by every instance of *A*. Suppose, however, that not every instance of *Y* are followed by every instance of *A* whereas every instance of *X* are followed by every instance of *A* in a possible world *W'* (suppose *W'* shares all the natural laws with our actual world). In this case, we have to conclude that '*A* causes *X*' is true, but '*A* causes *Y*' is false *in the actual world*. The point is, if a causal connection obtains or not cannot be determined solely by means of the actual occurrences or correlations. We need to have recourse to possible situations even to determine if a causal connection obtains in this actual world. And the same is true for functions. This is a notable character of modal concepts. Therefore, even to determine the actual teleological functions of the perception mechanism and motion mechanism of frogs, we need to look at possible situations where 'a flying insects' and 'a small moving black piece' have different extension. If we look at those possible situations, we can easily conclude that perceiving and catching a small moving black piece does not give nutrition to frogs, therefore 'perceiving and catching small moving black pieces' is not a proper teleological function of the perceptual mechanism and the motion mechanism of frogs.

As I see it, Fodor's objection can be overcome as argued above. However, *even if* Fodor's argument be right, the situation might not be so bad for the teleological approach. Suppose we accept Fodor's argument. Suppose that, confining the environment to the actual one where all the moving small black objects are flying insects, making frogs to catch moving small black pieces is nothing more or less than making frogs to catch flying insects. The teleological function of the frog's motion mechanism to make frogs to catch moving small black pieces is identical with the teleological function to make frogs to catch flying insects, and so is the perceptual mechanism. As both the motion mechanism and the perceptual mechanism are uniquely fixed, there is no indeterminacy of the teleological functions. However, although the teleological function is uniquely fixed, the content of *N* might not be uniquely fixed. *N* can be considered to represent that there is a moving small black piece, and can be

considered to represent that there is a flying insect as well. To see why, it is important to notice how a perception is formed by the frog's perceptual mechanism and how it is used by the frog's motion mechanism. Let us see, at first, the *perceptual* mechanism. *N* is produced by the frog's perceptual mechanism through the processing of the information such as moving, small, or black. It is not produced through the processing of the information such as flying, small or noisy. Therefore, it seems that *N* represents not that there is a flying insect but that there is a moving small black piece. We must, however, consider not only the perceptual mechanism but also the *motion* mechanism. The frog's neural state *N* is produced through the processing of the information such as moving, small and black. And it is used when the frog's motion mechanism makes the frog to stick out its tongue and try to catch an object. Does this *N*-using-process reflect the way *N* is produced? Probably not. It seems that the way *N* is used would not change even if the way *N* is produced changes. For example, even if the frog's perceptual mechanism changes and *N* were produced in a different way (e.g. produced through the processing of the information such as flying, small or noisy), *N* would be used in the same way: the frog's motion mechanism would make the frog to stick out its tongue and try to catch the object. Therefore, the way *N* is used does not reflect the way *N* is produced. Even if *N* is produced through the processing of the information such as moving, small and black, *N* does not represent that there is a moving small black piece, considering the way *N* is used. On the other hand, we cannot say that *N* represents that there is a flying insect, because we cannot find, in the frog's perceptual mechanism or motion mechanism, any factor that support the conclusion that *N* represents the content about insects. Do we have to conclude, then, that our intuition that *N* represents that there is a flying insect cannot be explained even by the teleological approach? Here we might be able to insist that the problem is in *our intuition itself*: we could doubt our intuition that the content of *N* is about insects. When we consider the perception of the frogs, we try to see how the frogs perceive objects, placing ourselves in the frog's situation. We put ourselves in the frog's place and consider the perceptual content. Therefore, it is not frog's perception but *our* perception. Human beings would be able to have both a perception that there is a flying insect in front and a perception that there is a moving small black piece in front, and would be able to catch a flying insect only when there is indeed a flying insect in front. However, if we consider the frog's

perception from the frog' point of view, we don't have any reason why we should think that *N* represents not that there is a moving small black piece but that there is a flying insect. The intuition that *N* represents a flying insect might not be a right intuition which is to be explained but a wrong intuition which should be discarded. What, then, does *N* represent? We could insist that *N* represents that there is something that is neither a flying insect nor a moving small black piece; the content of *N* is something that has not been finely differentiated. Understood in this way, the problem of the indeterminacy of the content might be resolved.[169]

A Problem for the Teleological Approach

So far, we have seen that intentionality has normative character and that only the teleological approach has a prospect of explaining normativity in the physicalistic framework. We have also seen that the problem of the indeterminacy of the content can be solved or dispelled in some sense by the teleological approach. In this section, however, I will deal with a serious objection to teleological approach, which throws a premise of teleological approach itself into doubt: it impeaches upon its historicism.

Donald Davidson's 'Swampman' thought experiment will make the point clear. 'Swampman' is what is born accidentally when thunder strikes a swamp. He is physically indistinguishable from Davidson himself, therefore behaves exactly like Davidson.[170] However, if intentionality essentially depends on history, then he lacks intentionality as he lacks his history. Therefore, even if Swampman has exactly the same brain state at the time when Davidson has when he thinks about something, Swampman does not think anything at all.

The Swampman problem poses a serious problem for the teleological approach. It questions the premise of the teleological approach, that is, the premise that functions are essentially determined by historical facts. It would be clear if we suppose the following situation. Suppose Swampman, after his birth, continues to live a happy life, and someday he finds a partner and has children of his own. These children grow up and have their own children in course of time. According to the teleological approach,

[169] I am helped in the argument of this paragraph by Nobuhara (1999).
[170] See Davidson (1987).

the organs of Swampman's descendants have functions, because those organs contribute to the owners' survival and existence. However, the teleological approach must insist that the organs of Swampman himself do not have any function – they are not really biological organs. But from the physical point of view, they are the same; Swampman's heart has exactly the same constitution as those of his descendants.[171]

We should notice here that the teleological approach has an externalistic factor. Facing the Swampman thought experiment, some of the externalists insist that Swampman's thoughts lack contents because Swampman lacks appropriate causal connection to objects and events in the world. The teleological approach might have to deny, in nearly the same manner as the externalism about mental contents does, that Swampman has mental contents like ours. Now the question is this: has the teleological approach offered enough metaphysical ground for asserting that Swampman really lacks proper functions or mental contents? At least, we can say that Millikan relies on a kind of causal *explanation* of the existence of organs, which might suggest that the teleological approach gives only epistemological implications, not metaphysical ones. At this stage, there, I myself am fairly close to be persuaded by John Heil's view expressed in the following quotation:

> … Swampman is dispositionally indistinguishable from Davidson. This, I suggest, is enough to endow Swampman's thoughts with significance. …
>
> I like to think of Swampman as a counter-example to externalism: if, on externalist grounds, we would be obliged to deny that Swampman has endless thoughts, externalism is mistaken.[172]

We must discard the teleological approach as a *metaphysical* theory of representations unless we get more metaphysical grounds than offered so far.

I must admit that Swampman poses a serious problem for the teleological approach. Among the philosophers who are interested in the problem of intentionality, it is widely accepted that normativity is an essential character of intentionality. We also have assumed this so far. However, the Swampman problem might suggest that we

---

[171] Thanks to Maeda (2003) for pointing out this possibility.
[172] Heil (2003), p.215.

should question this assumption. Dretske recently argues that although the biological functions that support normativity rely on historical facts, we should doubt if normativity is an essential character of intentionality.[173]

What Has the Teleological Approach Achieved?

What do we have to say about the teleological approach? Does it fail and have to be thrown away? First, we should notice that the three theories – the causal theory, the functional theory, and the teleological theory – are not opposed to each other: they are, in fact, compatible. We do not have to think that all the aspects of intentionality are explained by a single theory. Although we might have to discard metaphysical assertions of teleological approach, we will still be able to rely on it as an epistemological explanation of intentionality.

Second, even if the normativity might not be an *essential* character of intentionality, it is still true that the normativity would be one of the most important aspects of intentionality. And we don't have any prospective theory to explain normativity other than the teleological theory. The causal theory and the functional theory do not help in this matter at all: they are no better than the teleological theory with respect to explaining the normativity.

Third, and most importantly, the teleological approach has many advantages of its own. As we have seen above, when we think about the problem of what is a minimal requirement (necessary conditions) of intentionality, the teleological theory gives us good insight. It also helps us to understand how it is possible for intentionality to evolve in the history of living organisms; it tells us how a simple organism could have a (simple) proper function.

The teleological approach is still a very prospective one towards the understanding of intentionality and normativity in a physicalistic framework. It will give us more understanding about the concept of intentionality and normativity. When we think about the problem how to accommodate intentionality in a physicalistic framework, we can never ignore the insight given by the teleological approach.

---

[173] See Dretske (2001).

In this chapter, I have traced and developed several achievements in the theories of consciousness and intentionality. My general approach is to analyse consciousness by means of intentionality and to analyse intentionality physcalistically. This approach is very close to an eliminativism about consciousness and qualia. But notice that the approach does not just eliminate consciousness and qualia easily. The approach includes some attempts of *revision* of the concept of consciousness and qualia. As the sciences of consciousness have just started recently, there should exist much conceptual and empirical confusion. But if we try to find out how our minds are accommodated in a physical world, we have both to wait for the development of scientific research and to revise our ordinary concepts of consciousness and intentionality at the same time. What I did in this long chapter is to trace a prospective approach and see what have been achieved by it and what have not been achieved yet.

As to the problem of mental causation, the problem was generated from a very basic metaphysical background: in fact, it occurs, as we saw in the previous part of this thesis, when we presuppose the *minimal* requirements of physicalism. Therefore, in the case of mental causation, we should be able to apply the metaphysical achievements in the general theory of properties directly to the problem concerning mental causation, and so I did. As to the problem of consciousness and qualia, however, we are not in the same situation as the problem of mental causation. In the case of conscious and qualia, what we can do at the present is to apply the achievements of the metaphysics of properties in somewhat *in*direct way. I will attempt this in the next chapter.

11 The Problem of Qualia and the General Theories of Properties

In the previous chapter, I discussed the problem of consciousness (qualia) which has been bothering physicalism. I surveyed several attempts to accommodate consciousness within a physicalisitic worldview, and argued that the representation theory of consciousness is a very prospective one for physicalism although it still has many problems to be solved. In this chapter, I will show how we can look at the problem of consciousness and qualia from the perspective of *the general theory of properties* which I expounded in Chapter 6 and 7.

My goal in this chapter is to locate (or, at least, to suggest a prospect of how to locate) the problem of phenomenal qualia in philosophy of mind within a more general and wider problem of metaphysics – the general theories of properties[174]. Before going to the main task, let us see that qualia are considered, by many philosophers of mind, as *properties* of some sort: qualia are considered to be ontological entities which characterize particular objects or particular regions of space-time. When we experience qualia, we always experience them as whatever characterize particular objects or particular regions of space-time. When I experience a quale of blueness, I experience it as a property which characterizes, say, a part of the sky above me. When I experience a quale of pain, I experience it as a property which characterizes a part of my body. It is true that those properties might not be instantiated by actual particulars in this space-time. When, for example, I am dreaming of a green apple, the quale of greenness does not characterize an actual apple in this actual space-time. Still, qualia are considered to be properties of some kind.

According to the representation theory advocated in the previous chapter, qualia are not the properties of conscious state; they are the properties of intentional objects. Nonetheless, qualia are considered to be properties of some sort. Also, it might come

[174] Maeda (2009) draws a parallel between the theories of properties and the qualia problem, and suggests a general plan to deal with the qualia problem in view of the theories of properties. My argument in this chapter could be viewed as an attempt to develop the general plan according to the framework set in the previous chapters.

out that our intuition always miss-locate qualia: if a blue quale of the sky above me is nothing more than a brain state of me as some physicalists insist, then the quale is not really the property of the sky above me. However, we are still able to consider qualia as a kind of properties. This is to be confirmed before we go on and try to apply the general theories of properties to the problem of consciousness and qualia.

But what kind of properties? Here the metaphysics of properties kicks in.

What kind of properties ?

Qualia can be considered as properties of some sort. But what kind of properties? Here I list three presuppositions on qualia many of the current philosophers of mind seem to take.

(1) Intrinsicness

It is quite natural to consider qualia as *intrinsic* properties. Roughly speaking, an intrinsic property is 'a property that a thing has (or lacks) regardless of what may be going on outside of itself'.[175] For a subject of an experience to decide if a particular quale is instantiated by a particular (or a particular space-time region) or not, it seems to be sufficient that the subject looks at or pay one's attention to the space-time region only.

(2) Non-Functionality

Qualia are also considered, by many philosophers of mind, to be properties which *can not be completely grasped by their functional characterizations (causal profiles)*. Remember the conceivability argument and the knowledge argument discussed in chapter 10. Both arguments point out that qualia evade the functional characterization. Even if we completely describe a system's functional (or physical) aspect, there always seem to be a possibility that the system differs in its qualia: this is the basic intuition which both arguments rely on or point out.

[175] I borrow this expression from Yablo (1999). To use David Lewis's definition, an intrinsic property is a property shared by a particular and its *duplicate*. See Lewis (1983a) for his definition of extrinsic and intrinsic properties.

Now, it is very important, as I see it, to bear in mind that (1) Intrinsicness does *not* entail (2) Non-Functionality. A property being extrinsic or intrinsic does not have direct connection with a property being functional or not. In fact, I defended a version of Causal Theory of properties, in Chapter 6 and 7, which insists that a disposition (power) has causal profiles necessarily but it is nonetheless *intrinsic*.

Although (1) does not entail (2), (2) will entail:

(3) Quiddity

The second presupposition, Non-Functionality, seems to suggest that qualia are properties with quiddity. This will be discussed from the next section.

Now, I do not insist that qualia must, in fact, be intrinsic, non-functional, and quiddistic properties. I rather insist that many philosophers of mind, especially physicalists, whose arguments we discussed in the previous chapters, rely on an intuition about qualia, and that the intuition should presuppose some metaphysical views on properties. In fact, according to the discussions in the previous chapters, the second presupposition and the third presupposition are quite doubtful, although we may accept the first one.

In the following, I will show how the general theories of properties cast light on the problem of qualia.

The Problem of Qualia and the General Theories of Properties

As I suggested in the previous section, we can see metaphysical problems concerning general (not just mental) properties in the presuppositions about qualia problem. In fact, as I see it, the problem of qualia can be interpreted as a branch of the metaphysical problem concerning general properties. In Chapter 6 and 7, I examined several theories of properties and defended a version of causal theory. Remember that there are three main theories of properties: Humean Theory (Humean Categoricalism), Causal Theory (Weak version and Strong version), and the Identity Theory. As I discussed in Chapter 7, one of the main battle lines between Humean Theory and Causal Theory (and the Identity Theory) concerns the problem of *quiddity*. Humean Theory accepts the quiddity

of categorical properties, whereas Strong Causal Theory rejects it. Now the point mentioned in the previous section – qualia cannot be completely grasped by their causal profiles – seems to show that qualia might be considered as *properties with quiddity*.

Before going on, I have to point out that there is another theory of properties – Dualism. Dualism insists that there are two kinds of properties – dispositions and categorical properties, and that one is not reducible to the other.[176] Although I cannot take up this in detail here, I would like to give just one comment. As I see it, the relation between properties and their causal profiles should be common to *all* kinds of properties. If there are two kinds of properties, one of which being such that the properties have their causal profile contingently (i.e. categorical properties) and the other of which being such that the properties have their causal profile necessarily (i.e. dispositions), then we face a question: why are there two kinds properties at all? What explains the difference between two kinds of properties? If we have a theory of properties which analyzes all properties in the same way, this is, I believe, much more ideal as a philosophical theory.

If the relation between properties and their causal profiles are common to all kinds of properties (as I believe they should be), then how we think about the relation between properties and their causal profiles should have something to do with how we deal with the problem of qualia. In the following, I will keep myself neutral about the theories of properties, for the time being, and see what consequences we get about the problem of qualia from the theories of properties.

Humean Theory of Properties and Qualia

Let us start with Humean Theory of Properties. As we saw in Chapter 6 and 7, Humean Theory accepts quiddity of properties in general. Therefore, according to Humean Theory, it is possible to insist that both physical properties and mental properties (especially qualia) are in the same situation as to the relation to their causal profiles: properties of both kinds have contingent relation to their causal profiles. A physical property, say a molecular structure, has a disposition, say being fragile, in this actual world; but the same physical property has a different disposition, say being sturdy, in a

[176] See, for example, Place (1996) and Molnar (2003) for Dualism of properties.

possible world. In the same way, a mental state, say feeling pain, has such and such function (causal profile); but the same mental state has another function in a possible world. Suppose Humean Theory is correct. What consequences does it have to the problem of qualia?

Firstly, as mentioned above, the fact that qualia seem to have qualitative aspects that cannot be exhausted by their causal profiles, does not pose a problem *just for physicalism* any more. That is because there is no difference between physical properties and qualia in that both properties cannot be exhausted by their causal profiles.

Secondly, the problem of how qualia could be identical with or supervene on physical properties, the problem we discussed in the previous chapters, might come out not to be a problem *just for physicalism* any more. That is because the relation between properties and their causal profiles is contingent, and perhaps could be seen as brute facts which need not be explained, *regardless of the fact* that they are physical properties or non-physical properties – qualia. In this case, the remaining problem might be to find out how we could connect qualitative aspects of qualia with quiddities of physical properties.

Thirdly, it is sometimes argued in the philosophy of mind that qualia could be epiphenomenal, but Humean Theory of properties has the same problem of epiphenomenalism as well, as Robinson points out.[177] If two properties could be different with each other only in their quiddities with their causal profiles being identical, then it is quite plausible to think that their quiddities are causally idle. It is true that there are many problems concerning quiddity (as I discussed in Chapter 7), but the problem might not be the problem concerning qualia in particular. The problem of epiphenomenalism comes out to be not a problem solely for qualia but a problem for properties in general.[178]

All in all, if we take Humean Theory of properties, the problem of qualia could be subsumed under a branch of general theory of properties. We might be able to say that the problem of qualia (the problem particularly concerning qualia) *as a problem in the philosophy of mind* is eliminated by means of generalizing it as a more basic ontological problem. This, however, is not to say there remain only easy task. Rather

[177] Cf. Robinson (1993).
[178] Cf. Maeda (2009), p.34.

philosophers face a very difficult problem to explain or establish the relation between quiddity and powers, as, according to Humean Theory, quiddity is considered to be such ontological items as freely floating from their causal profiles. We need to find out a plausible relation between them, otherwise quiddity remains as a deep mystery.

Causal Theory of Properties and Qualia

Now let us consider the Causal Theory of Properties. We have two versions of Causal Theory, Strong Causal Theory and Weak Causal Theory. As I showed in Chapter 6, the Weak Causal Theory incorporates quiddity and, therefore, is in exactly the same situation as Humean Theory discussed in the previous section. I will, therefore, focus on the Strong Causal Theory here.

Causal Theory simply rejects quiddity. Therefore, it cannot so quickly take the same line as Humean Theory which tries to deal with phenomenal qualia and other (physical) properties in the same manner. Obviously, there are two ways to take: accept qualia and try to find a connection between qualia and physical properties, or simply reject qualia. If we take the former, the problem of qualia is, for the Causal Theory, a specific problem of the philosophy of mind: the problem of qualia cannot be subsumed under a more general problem in the ontology of properties. Here again, philosophers face a great stumbling block. Suppose one takes functionalism. Suppose also one takes (2) Non-Functionality, and thereby accept (3) Quiddity. This view obviously faces the conceivability argument and the knowledge argument. This is, in fact, the situation where functionalism in the philosophy of mind has been. As to mental states like belief and desire, functionalism work well. This is because those intentional states could be regarded as the properties exhausted by causal profiles. On the other hand, as to the mental properties which are not exhausted by causal profiles (or which seem not to be exhausted by causal profiles), functionalism does not work well. The conceivability argument and the knowledge argument in the previous chapter well indicate this point.

Therefore, the most natural and plausible approach to the problem of qualia for the proponents of Causal Theory to take is to take the second way, that is, to reject (i.e. to eliminate) qualia, if, of course, we have a prospect to do so properly within the framework of physicalism. This is why the previous chapter was devoted to the

representation theory of consciousness. The representation theory of consciousness, in a way, eliminates qualia, but in doing so, the theory tries to revise the concept of consciousness (and intentionality). Although there are many problems to be solved about the representation theory of consciousness (and the teleological approach to intentionality), I have, at least, showed that it is prospective. And more importantly, with a prospective theory of properly eliminating qualia in a physicalistic framework at hand, the Causal Theory of properties would also gain its plausibility, which is what I have been trying to do in the whole thesis.

Identity Theory and Qualia

Finally, let me briefly comment on Identity Theory. As we discussed in Chapter 6, Identity Theory has it that there is only one kind of properties but there are two aspects of it – powers and quality. What, then, would be the consequence on the problem of qualia, if we take Identity Theory as a theory of general properties.

If we take Identity Theory, the problem of qualia could be subsumed under a branch of the general theory of properties, as there is only one kind of properties regardless of it is physical or mental. This is similar to the case of Humean Theory discussed above, although Identity Theory does not accept quiddity as ontological items freely floating from their causal profiles. Unlike Humean Theory, Identity Theory has at hand a possible solution to the problem of finding out the relation between quiddity and causal profile – *identity*. However, this does not immediately mean that Identity Theory is free from serious problems. As we saw in chapter 6, John Heil's identity theory is presented as 'a default view' which avoids the difficulties of both pure dispositionalism and pure quality theory (categoricalism). But, as I also showed it, the theory leaves the *unintelligibility of the identity relation* aside. I discussed that Heil's view faces the difficult problem of how to understand the identity between powers and qualities. I also insisted that his example of Necker Cube is not persuasive enough.

There is a similarity between the difficulty which Heil's identity theory faces and the difficulty which the mind-body (or mind-brain) identity theory in the philosophy of mind faces. Both theories have to find out how to identify quality with their causal profiles. Both theories have to explain, in some way or others, how the

properties which seem to be so different with each other could be identical. There should be much work to be done, and I must say that the work has not been done enough compared with the causal theory of properties.

Concluding Remarks

The problem of consciousness (especially, of qualia) has been worrying physicalists. As many philosophers see it, qualia (or phenomenal properties) seem to be something more than mere functional properties. The problem has been considered to be the problem arising from the physicalist presupposition. As I showed in this chapter, however, the problem is very likely to be a problem concerning QUIDDITY, a problem in the general theory of properties. The problem of qualia for physicalists is, in fact, a form of the problem concerning the relation between the essential intrinsic qualities (quiddity) of a property and its causal powers (causal profiles). Which view in the general theory of properties we should take affects which view in the study of qualia we should take. If we take a causal theory of properties (as I took in the previous part of the thesis), then we are very likely led to an eliminativist view on qualia. I have presented the representation theory of consciousness and the teleological approach to intentionality as an example.

As to the general theory of properties, I expounded my own view, Causal Trope Theory in comparison with other views. With this theory in mind, I concluded that we could find a way to understand mental causation in the direction which Shoemaker suggested, although there still remain many problems to be solved. As to the problem of consciousness, I defended and developed a representation theory of consciousness independent of the general theory of properties. And then, I viewed the problem of consciousness from the perspective of the general theories of properties and showed that my own view on properties and a representation theory of consciousness are good combination. The arguments I presented for a version of the causal theory of properties and for a version of the representation theory of consciousness are not conclusive, of course. However, if I have shown that there is a prospect that the problem of mental causation and consciousness will be subsumed into the general theory of properties and that there are many philosophical merits for us to view those problems in this way, the

purpose of this thesis is accomplished.

Appendix: The Problem of Extrinsicness of Mental Properties

Roughly speaking, there are three problems about mental causation: the problem from Anomalous Monism, the problem from causal exclusion, and the problem of the extrinsicness of mental properties. We have already seen the first two problems in the thesis. In this appendix, I will very briefly deal with the third problem: the problem of the extrinsicness of mental properties.

Extrinsicness of Content Properties

There are many mental properties. Some of them are 'content properties' – properties with propositional contents. For example, the property of believing that the earth is flat and the property of desiring that I should drink a glass of water are content properties (their propositional contents being 'the earth is flat' and 'I should drink a glass of water' respectively).

Since Putnam offered a famous thought experiment on twin earth, it is widely considered that content properties are not intrinsic but *extrinsic* properties of the subject. That is to say, they are considered to be the properties that the subject has in virtue of its having some relations with other objects. Putnam's twin earth thought experiment is as follows. Suppose there is a twin earth which is the exact duplicate of our earth except just one difference: in twin earth there are *XYZ*, instead of H_2O, that look exactly like water in our earth. On the twin earth is twin-you that cannot be distinguished from you with respect to the intrinsic properties (suppose for the sake of argument that your body does not contain H_2O). Suppose that you desire that you should drink a glass of water. Then, at the very same time on the twin earth, twin-you should have that kind of desire, for you and twin-you are in exactly the same brain state. You and twin-you, nonetheless, have different mental contents with each other, for your desire is about H_2O whereas the desire of twin-you is about *XYZ*. To decide what content property a subject has at a time,

it must be taken into account what kind of external environment the subject is in.[179]

Content properties, thus, are considered to be extrinsic properties. However, if the intrinsic properties of two subjects are completely equal, their physical behaviours should be completely equal as well. For example, if you and twin-you share exactly the same intrinsic properties, you and twin-you always respond to the same questions in the same manner. It seems, then, that a subject's physical behaviours are completely determined by its intrinsic properties.

Here kicks in a question. If a subject's behaviours is completely determined by the subject's intrinsic properties and content properties are extrinsic as many philosophers believe them to be, then what would content properties of the subject do in determining the subject behaviours? What is left for content properties to do in causing the subject's behavioural outputs? It seems that content properties, which are typical mental properties, have *no* causal efficacy in bringing about the physical or bodily behaviours of the subject, which seems to show that mental causation is impossible as far as content properties go. This is the problem of the extrinsicness of mental properties (content properties).[180]

The Problem of Extrinsicness and Physicalism

This problem, however, is not the main theme of the thesis. That is because while the exclusion problem is a problem for physicalism, the problem of extrinsicness is not a problem for physicalism alone. The exclusion problem is a problem for physicalism because it depends on the principle of supervenience and the principle of causal closure which are necessary conditions for a minimal physicalism (see Chapter 2 and Section 4-1). The problem of extrinsicness, on the other hand, has nothing to do with whether you take physicalism or not. Suppose you are a Cartesian dualist. You insist that mind and body are different substances; a non-physical (mental) event causes and is caused by a physical event; neither the principle of supervenience nor the principle of physical closure obtain. However, if content properties are extrinsic properties, as Putnam and other philosophers insists, then you will be troubled with the problem of extrinsicness as

[179] See Putnam (1975a), "The meaning of 'meaning'".
[180] Cf. Kim (1998) pp.35-37.

long as you admit that only intrinsic properties of the subject have causal efficacy in bringing about physical (bodily) behaviours. In this way, even a robust Cartesian dualist would be troubled with the problem of extrinsicness. Therefore, it is not a problem solely for physicalism.[181]

How Can We Solve the Problem of Extrinsicness?

The main theme of the thesis has been to examine the problem of mental causation in a physicalistic framework. Although the problem of extrinsicness is not a problem for physicalism alone and it does not concern primarily the present thesis, I will just briefly suggest how this problem could be solved.

Very briefly put, my suggestion is to *reject* externalism about mental contents and to take internalism instead. As I see it, we should cast doubt on the premise of the extrinsicness problem itself: the premise that content properties are extrinsic properties. The debate between externalism and internalism of content properties is not a problem that we can fully examine in just one chapter or two. Here I am going to offer some suggestions that might help the problem of extrinsicness to be solved.

It is true that content properties are described in terms of propositional contents. And it is probably true that a content property of you at a time and the corresponding content property of twin-you at the corresponding time are described in terms of different propositional contents; 'I should drink a glass of H_2O' in case of you and 'I should drink a glass of *XYZ*' in case of twin-you. However, it does not immediately follow from this that you and twin-you have different properties. My daughter is described as 'my daughter'; she is also describes as 'my father's granddaughter'. But this does not mean that my daughter has two different properties corresponding to two different descriptions.

The point is: do we really have to pose different content properties for each different description? As I suggest it, we do not have to, if we ascribe a *disposition D*, 'to drink H_2O in our actual world *and* to drink *XYZ* in twin earth', both to you and twin-you. Suppose twin-you were brought from the kitchen of twin-you's house on twin earth to the kitchen of your house on our earth, instantly, at the time twin-you has a

[181] Cf. Kim (1998) and Mino (2004).

desire to drink a glass of water (*XYZ*). How would he/she behave in your kitchen? He/She would drink a glass of H_2O (that is, 'water' in our world) in your kitchen on the earth because he/she has a disposition *D*. Conversely, if you were brought to twin earth, you would drink a glass of *XYZ* (that is, 'water' in twin earth) because you have a disposition *D*.[182] It seems that if we introduce a disposition, we do not have to pose different content properties for each descriptions. This might suggest that there really do not exist content properties. Or this might suggest that content properties of you and twin-you are really identical with each other even though they are differently described in terms of different propositional contents.[183]

Notice that dispositions are (plausibly) considered to be intrinsic properties and causally efficacious as I discussed in Chapter 6 and 7. Therefore, we do not have to be bothered with the problem of the extrinsicness any more if we identify content properties with dispositions. All in all, as in the 'Swampman Case' discussed in Chapter 10, I would like to take the problem of the extrinsicness of mental properties as a counter-example to externalism until a more persuasive argument for supporting externalism is offered.

[182] A dart-tossing model offered by John Heil might help. See Heil (2003) pp.210-211.
[183] For a general plan of identifying content properties with dispositions, see Crane (1998b) and Heil (2003). Mino (2004) suggests a similar view although it does not explicitly introduce dispositions as a candidate for content properties.

Bibliography

Achinstein, P. (1974). The identity of properties. *American Philosophical Quarterly* 11: 257-75.

Armstrong, D. M. (1968). *A Materialist Theory of Mind*. London: Routledge & Kegan Paul.

Armstrong, D. M. (1983). *What is a Law of Nature?* Cambridge: Cambridge University Press.

Armstrong, D. M. (1989). *Universals: An Opinionated Introduction*. Westview Press.

Armstrong, D. M., Place, U. T., and Martin, C. B. (1996). *Dispositions: A Debate*. London: Routledge.

Baker, L. R. (1993). Metaphysics and mental causation. in Heil and Mele eds. (1993).

Bird, A. (2007). *Nature's Metaphysics: Laws and Properties*. Oxford: Clarendon Press.

Block, N. (1986). Advertisement for a semantic for Psychology. *Midwest Studies in Philosophy* 10: 615-678.

Block, N. (1995). On the confusion about a function of consciousness. *Behavioral and Brain Sciences* 18: 1-41.

Bogdan, R. ed. (1986). *Belief*. Oxford: Oxford University Press.

Buller, D. J. (1998). Etiological theories of functions: a geographical survey. *Biology and Philosophy* 13: 505-27.

Chakravartty, A. (2005). Causal realism: events and processes. *Erkentnis* 63: 7-31.

Chalmers, D. (1997). *The Conscious Mind: In Search of a Fundamental Theory*: Oxford University Press.

Churchland, P. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy* 78, Reprinted in Lycan ed. (1990).

Clapp, H. (2001). Disjunctive properties: multiple realizations. *The Journal of Philosophy* 98: 113-36.

Crane, T. ed. (1996). *Dispositions: A Debate by D. M. Armstrong, C. B. Martin and U. T. Place*. London and New York: Routledge.

Crane, T. (1998a). Intentionality as the Mark of the Mental. In A. O'Hear (ed.) *Current*

Issues in Philosophy of Mind. Cambridge: Cambridge University Press.

Crane, T. (1998b). The efficacy of content. In *Human Action, Deliberation and Causation*, edited by J. Bransen and S. E. Cuypers: 199-233.

Crane, T. (2001). *Elements of Mind*. Oxford: Oxford University Press.

Crane, T. (2003). *The Mechanical Mind: Second Edition*. London: Routledge

Davidson, D. (1969). The individuation of events. reprinted in Davidson (1980).

Davidson, D. (1970). Mental events. Reprinted in Davidson (1980).

Davidson, D. (1980). *Essays on Actions and Events*. Oxford: Oxford University Press.

Davidson, D. (1987). Knowing one's own mind. reprinted in Davidson (2001)

Davidson, D. (1993). Thinking causes. In Heil and Mele eds. (1993).

Davidson, D. (2001). *Subjective, Intersubjective, Objective*. Oxford: Oxford University Press.

Davies, P. M. (1994). Trouble for direct proper functions, *Nous* 28.

Dennett, D. C. (1991). *Consciousness Explained*. Boston: Little Brown.

Dretske, F. (1986). Misrepresentation. in R. Bogdan ed. (1986)

Dretske, F. (2001). Norms, history and the mental. In *Naturalism, Evolution and Mind*, edited by D. M. Walsh, Cambridge: Cambridge University Press

Ehring, D. (1997). *Causation and Persistence: A Theory of Causation*. Oxford University Press.

Ehring, D. (2003). Part-whole physicalism and mental causation. *Synthese* 136: 359-88.

Elder, C. L. (1998). What versus how in naturally selected representations. *Mind* 107:349-63

Ellis, B. (2002). *The Philosophy of Nature: A Guide to the New Essentialism*. Chesham: Acumen.

Fara, M. (2006). Dispositions. *Stanford Encyclopedia of philosophy*, http://plato.stanford.edu/entries/dispositions/

Fodor, J. (1974). Special Sciences, or the Disunity of Science as a Working Hypothesis. *Synthese* 28: 97-115.

Fodor, J. (1990). *A Theory of Content*. Cambridge, MA: MIT Press.

Gibb, S. C. (2004). The problem of mental causation and the nature of properties. *Australasian Journal of Philosophy* 82: 464-76.

Gillett, C. (2002). The Dimensions of Realization: A Critique of the Standard View.

Analysis 62: 316-323.

Gillett, C. (2003). The Metaphysics of Realization, Multiple Realizability, and the Special Science. *The Journal of Philosophy*: 591-603.

Gillett, Carl and Bradley Rives (2005). The non-existence of determinables: or, a world of absolute determinates as default hypothesis. *Nous* 39: 483-504.

Hawthorne, J. (2006). Causal structuralism. In *Metaphysical Essays*, Oxford: Oxford University Press.

Harman, G. (1990). The intrinsic quality of experience. *Philosophical Perspectives* 4.

Heil, J. (2003). Multiply Realized Properties. In Walter and Heckmann eds., *Physicalism and Mental Causation*, Imprint Academic

Heil, J. (2003). *From an Ontological Point of View.* Oxford: Oxford University Press.

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

Honderich, T. (1982). The argument for anomalous monism. *Analysis* 42.

Horgan, T. (1997). Kim on mental causation and causal exclusion. *Philosophical Perspectives* 11:165-84.

Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly* 32: 127-138.

Jackson, F. (1998). *From Metaphysics to Ethics*: Oxford University Press.

Jacob, P. (1997). *What Minds Can Do*. Cambridge: Cambridge University Press.

Kim, J. (1976). Events as property exemplifications. reprinted in Kim (1993b).

Kim, J. (1984a). Concepts of supervenience. *Philosophy and Phenomenological Research* 45. Reprinted in Kim (1993b).

Kim, J. (1984b). Self-understanding and rationalizing explanations. *Philosophia Naturalis* 82.

Kim, J. (1987). 'Strong' and 'global' supervenience revisited. Reprinted in Kim (1993b).

Kim, J. (1989a). Mechanism, purpose, and explanatory exclusion. *Philosophical Perspectives* 3. Reprinted in Kim (1993b).

Kim, J. (1989b). The myth of nonreductive physicalism. *Proceedings and Addresses of the American Philosophical Association* 63. Reprinted in Kim (1993b).

Kim, J. (1992). Multiple realization and the metaphysics of reduction. *Philosophy and Phenomenological Research* 52. Reprinted in Kim (1993b).

Kim, J. (1993a). The nonreductivist's troubles with mental causation. In *Mental Causation*, edited by Heil and Mele. Reprinted in Kim (1993b).

Kim, J. (1993b). *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press.

Kim, J. (1996). *Philosophy of Mind*. Boulder, CO: Westview Press.

Kim, J. (1998). *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: MIT Press.

Kim, J. (2011). *Philosophy of Mind: Third Edition*: Westview Press.

LePore, E. and B. Loewer (1989). More on making mind matter. *Philosophical Topics* 17: 175-91.

Lewis, D. (1966). An argument for the identity theory. *Journal of Philosophy* 63. Reprinted in Lewis (1983c).

Lewis, D. (1983a). Extrinsic properties. Reprinted in Lewis (1999)

Lewis, D. (1983b). New work for a theory of universals. *Australasian Journal of Philosophy* 61. reprinted in Lewis (1999).

Lewis, D. (1983c). *Philosophical Papers Volume 1*. Oxford: Oxford University Press.

Lewis, D. (1986). *On the Plurality of Worlds*. Oxford: Blackwell.

Lewis, D. (1997). Finkish dispositions. *Philosophical Quarterly* 47. Reprinted in Lewis (1999).

Lewis, D. (1999). *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Loar, B. (1997). Phenomenal states (revised version). In *The Nature of Consciousness: Philosophical Debates*, edited by N. Block, O. Flanagan, and G. Guzeldere, 597-616. Cambridge, MA: MIT Press.

Loux, M. J. (2002). *Metaphysics: A Contemporary Introduction, Second Edition*. London: Routledge.

Lowe, E. J. (2000). Causal closure principles and emergentism. *Philosophy* 75: 571-585.

Lowe, E. J. (2001). Dispositions and laws. *Metaphysica* 2: 5-23.

Lowe, E. J. (2006). *The Four-Category Ontology: A Metaphysical Foundation for Natural Science*. Oxford: Clarendon.

Lycan, W. ed. (1990). *Mind and Cognition: A Reader*. Oxford: Basil Blackwell.

Lycan, W. (2001). The case for phenomenal externalism. *Philosophical Perspectives 15: Metaphysics*: 17-35.

Maeda, T. (2004). Intentionality and the Teleological Functionalism (*in Japanese*). in Nobuhara ed. (2004).

Maeda, T. (2009). Quidditistic Qualia (*in Japanese*). *Journal of the Japan Association for Philosophy of Science* 37: 29-38.

Marcus, E. (2005). Mental causation in a physical world. *Philosophical Studies* 122: 27-50.

Martin, C. B. (1994). Dispositions and conditionals. *Philosophical Quarterly* 44: 1-8.

McKitrick, J. (2003). The bare metaphysical possibility of bare dispositions. *Philosophy and Phenomenological Research* 66: pp.349-69.

McKitrick, J. (2003). A case for extrinsic dispositions. *Australasian Journal of Philosophy* 81: 155-74.

McKitrick, J. (2004). A defence of the causal efficacy of dispositions. *Sats: Nordic Journal of Philosophy* 5: 110-130.

McKitrick, J. (2005). Are dispositions causally relevant? *Synthese*.

McLaughlin, B. (2007). Mental causation and Shoemaker-Realization. *Erkenntnis* 67.

Mellor, D. H. (1974). In defence of dispositions. *Philosophical Review* 83: 157-81.

Mellor, D. H. (2000). The semantics and ontology of dispositions. *Mind* 109: 757-80.

Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.

Millikan, R. G. (1989). Biosemantics. *Journal of Philosophy* 86. reprinted in Millikan (1993).

Millikan, R. G. (1993). *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press.

Mino, T. (1995). Anomalous monism and mental causation (*in Japanese*). *Jinbun Kenkyu, Osaka City University* 47: 43-61.

Mino, T. (2004). Mental causation and physicalism (*in Japanese*). in Nobuhara ed. (2004).

Molnar, G. (2003). *Powers: A Study in Metaphysics*. Edited by S. Mumford. Oxford: Oxford University Press.

Mumford, S. (1998). *Dispositions*. Oxford: Oxford University Press.

Mumford, S. (2004). *Laws in Nature*. Routledge

Nagel, E. (1961). *The Structure of Science*. New York, NY: Harcourt, Brace & World.

Neander, K. (1995). Misrepresenting and malfunctioning. *Philosophical Studies* 79: 109-41.

Nobuhara, Y. (1999). *Contemporary Philosophy of Mind (in Japanese)*. Keiso Shobo.

Nobuhara, Y. ed. (2004). *Series Philosophy of Mind Vol.1* (*in Japanese*). Keiso Shobo.

Noordhof, P. (1998), Do tropes resolve the problem of mental causation? *Philosophical Quarterly* 48: 221-6.

Papineau, D. (1990). Why supervenience? *Analysis* 50.

Papineau, D. (1993). *Philosophical Naturalism*. Blackwell Publishing.

Pereboom, D. (2002). Robust nonreductive materialism. *The Journal of Philosophy* 99: 499-531.

Place, U. T. (1996). *Structural properties: categorical, dispositional or both?*. in Crane ed. (1996).

Poland, J. (1994). *Physicalism: The Philosophical Foundations*. Oxford: Clarendon Press.

Price, C. (2001). *Functions in Mind*. Oxford: Oxford University Press.

Prior, E. W., Pargetter, R. and Jackson, F. (1982). Three theses about dispositions. *American Philosophical Quarterly* 19: 251-7.

Prior, E. W. (1985). *Dispositions*. Aberdeen: Aberdeen University Press.

Putnam, H. (1975a). The meaning of 'meaning'. In *Language, Mind, and Knowledge: Minnesota Studies in the Philosophy of Science* 7, edited by K. Gunderson. Reprinted in Putnam (1975b).

Putnam, H. (1975b). *Mind, Language, and Reality*. Cambridge: Cambridge University Press.

Ramachandran, V. S. (1999). *Phantoms in the Brain: Probing the Mysteries of the Human Mind*: William Morrow Paperbacks.

Robb, D. (1997). The properties of mental causation. *Philosophical Quarterly* 47: 178-94.

Robb, D. (2001). Reply to Noordhof on mental causation. *Philosophical Quarterly* 51: 92-3.

Robinson, D. (1993). Epiphenomenalism, laws, and properties. *Philosophical Studies* 69: 1-34.

Rosenthal, D. (1986). Two concepts of consciousness. *Philosophical Studies* 49:

329-359.

Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson.

Shapiro, L. A. (2000). Multiple realizations. *The Journal of Philosophy* 97: 635-54.

Shoemaker, S. (1980). Causality and properties. Reprinted in Shoemaker (2003).

Shoemaker, S. (1998). Causal and metaphysical necessity. *Pacific Philosophical Quarterly* 79. Reprinted in Shoemaker (2003).

Shoemaker, S. (2001). Realization and mental causation. In *Physicalism and Its Discontents*, edited by C. Gillett and B. Loewer. Reprinted in Shoemaker (2003).

Shoemaker, S. (2003). *Identity, Causes, and Mind: Expanded Edition*. Oxford: Oxford University Press.

Shoemaker, S. (2007). *Physical Realization*. Oxford: Oxford University Press.

Sosa, E. (1984). Mind-body interaction and supervenient causation. *Midwest Studies in Philosophy* 9

Stoutland, F. (1980). Oblique causation and reasons for actions. *Synthese* 43.

Suzuki, T. (2003). Qualia and the hard problem of consciousness (*in Japanese*). in Nobuhara ed. (2003).

Tye, M. (1997). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind.* A Bradford Book.

Wilson, J. (1999). How superduper does a physicalist supervenience need to be? *Philosophical Quarterly* 49: 33-52.

Yablo, S. (1992). Mental causation. *The Philosophical Review* 101: 245-80.

Yablo, S. (1999). Intrinsicness. *Philosophical Topics* 26: 479-505.

Yablo, S. (2007). The seven habits of highly effective thinkers. http://mit.edu/~yablo/effthink.html.