Spring 2013

Emergence, Mental Causation, and Exclusion

Kyle Rindahl
University of Colorado Boulder

Emergence, Mental Causation, and Exclusion

Kyle Rindahl
Department of Philosophy
University of Colorado, Boulder

Thesis Advisor:
Dr. Robert Rupert | Department of Philosophy

Committee Members:
Dr. Robert Rupert | Department of Philosophy
Dr. Wesley Morriston | Department of Philosophy
Dr. Eliana Colunga | Department of Psychology

Defended: April 8th, 2013

Abstract

Emergence as a general concept has had an interesting and varied history. In this paper, I critically analyze various emergentist concepts within the context of philosophy of mind. In this portion of the project, I first analyze the history of emergentism, paying great debt to the contributions of a group of philosophers known as the British Emergentists. Building upon their foundation, I then highlight a few concepts of great interest for emergentism as a mind-body theory, and, from there, build a definition of an emergent property. I highlight *irreducibility* as the main concept for the emergentist. The irreducibility claim for the emergentist is made up of four separate claims: 1) emergent properties *supervene* on base properties with *nomological necessity*, but not *logical necessity*, 2) emergent properties are not *reductively explainable* in terms of their basal constituents, 3) the laws relating emergent properties with their basal properties are *fundamental, irreducible laws*, and 4) emergent properties have *causal powers of their own* and these properties are not reducible to the causal powers of their constituents.

Given this definition, I argue that emergentism is something much closer to *property dualism* as opposed to *physicalism*. This is a significant result, because it is generally thought that emergentism is one of the paradigmatic examples of a *nonreductive physicalist* theory. Thus, the emergentist must deal with the same problems as the property dualist, most importantly the problem of *mental causation*. To highlight this problem, I present Jaegwon Kim's *exclusion argument*. From there, I offer a possible solution on behalf of the emergentist, but argue that this solution does not solve the problem. Thus, the emergentist still faces the issue of explaining how an irreducible mental cause can fit into a physical world.

Table of Contents

Emergence, Mental Causation, and Exclusion

1. Introduction

The concept of "emergence" has had an interesting and varied history. It has enjoyed time in the spotlight of mainstream philosophy and has endured periods in the depths of intellectual distaste due to charges of mystical unacceptability. Despite the latter, emergence has found its way back into central discussions of disciplines including cognitive science, computer science, complexity studies, and academic philosophy- most importantly, philosophy of science, and philosophy of mind. The term 'emergence' has different implications relative to the context to which it is applied, thus the notion of emergence is still a rather inherently vague concept.

What is of interest here is emergence in the context of philosophy of mind. It is generally agreed that the history of emergence began with John Stuart Mill and was subsequently expanded on by philosophers such as Samuel Alexander and C.D. Broad among others. These three philosophers (among others) are collectively known as the British Emergentists. Many consider the period when this group was writing on the subject the heyday of emergentist theories. However, by the mid 20th century, emergentism was ushered into a bleak existence by the logical positivists, who dominated mainstream philosophy during this period. Their hyper-empiricist and anti-metaphysical doctrines left no room for "spooky" concepts such as emergence. However, during the 1970s growing dissatisfaction with reductionism with regard to mental phenomena led to the construction of nonreductive theories of the mental, and in turn, the re-emergence of emergentist concepts.

To state the emergentist moniker eloquently, one could say "the whole is more than the sum of its parts." The whole is indeed grounded in its parts, and there is no need for

things to be added from the outside such as the postulation of Cartesian souls. Despite this, the emergentist still maintains that the whole cannot be fully reduced to its constituent base conditions. Thus, the concept of *irreducibility* (the claim that mental properties and physical properties are distinct) plays a major role in the discussion of emergence. Even though emergent properties are irreducible to their basal components, many emergentists claim that the mental *supervenes* (the claim that the mental depends on the physical in some way) on the physical; thus there is a tight connection between the two. In this sense, emergentism is a potential middle ground between reductive physicalism and substance dualism- both of which are considered untenably extreme by many, but of course not all. Thus, emergentism is intuitively appealing to those who find reduction and substance dualism unappealing or unacceptable for whatever reason.

Like any nonreductive theory of mind, emergentism has to deal with the very difficult problem of mental causation. The problem of mental causation can be stated as such: if the world is fundamentally physical, how can nonphysical mental properties exert causal influence on physical properties? For mental properties to have any causal efficacy, it seems that this would violate the laws of physics. For any naturally inclined philosopher, this result is prima facie unacceptable. If mental properties do not have any causal powers, then it is not clear that mental properties exist in any meaningful sense, or at least not in the way we intuitively think they exist. The formulation of this dilemma is largely credited to Jaegwon Kim in his famed "exclusion argument." Both Kim and his argument will be of much importance in what follows.

The goal of this paper is two-fold. The first part of the goal is to make sense of the concept of emergence for the purpose of eventually applying it within the current debates in

philosophy of mind. This will require a few steps. First, in Section 2, I will present the historical development of emergence, paying close attention to the contributions made by the British Emergentists. In section 3, I will then build upon the theories of the British Emergentists and construct a definition of (hopefully) the necessary and sufficient conditions of an emergent property. This is an important task for a few reasons. First, as already mentioned, emergentism is an admittedly vague concept that can be applied to many contexts; consequently, it can have different theoretical implications depending on which sense of emergentism one applies. Second, it will lay the theoretical framework the emergentist will be allowed to work within when trying to solve the philosophical problems that I will present in later sections of the paper. Section 4 will then apply the definition of an emergent property to the context of philosophy of mind with the purpose of establishing its relationship to other mind-body theories. I will conclude that emergentism holds a distinctive and interesting place within the theoretical landscape.

The second part of the goal is to assess the tenability of the emergentist mind-body theory in the face of the problem of mental causation. In section 5, I will present the causal exclusion problem as presented by Kim. In section 6, I will then offer what I believe to be an essentially emergentist response that directly argues for the causal efficacy of emergent mental properties. I will then critically examine this response.

2. The Rise, Fall, and Re-Emergence of Emergence

The history of emergentism has been argued to go all the way back to the classical period with contributions from such philosophers as Galen and Aristotle, but for all intents and purposes, the real history of emergentism began with "British Emergentism." I say "real" history, because it was this group that was responsible for the development of important emergentist themes that are still being used in academic setting today. British Emergentism began around the late 19th century, beginning with J.S. Mill, and faded around the mid 20th century. This may seem like a fairly small period of livelihood but I will explore some possible explanations for its quick departure at the end of this section. During this part of the section, I will examine what, if anything, its period of distaste means for the philosophical significance of the theory of emergence.

The heyday of emergentism was in the 1920's. Three major emergentists all published their main works within this decade: Samuel Alexander's *Space, Time and Deity* (1920), C. Lloyd Morgan's *Emergent Evolution*, (1923) and C.D. Broad's *The Mind and its Place in Nature* (1925). There were of course other emergentists that came after this tradition, most notably Roy Wood Sellars (1922), Ernest Nagel (1961), Karl Popper (1977) and Mario Bunge (1977). While this latter group of philosophers that falls outside the British Emergentist tradition were also influential in the development of theories of emergence, to fully discuss them in the proper context would require a study all its own. That being said, I will choose to focus on the thought of J.S. Mill, S. Alexander, and C.D. Broad; more specifically, I will focus on their works *A System of Logic* (1843), *Space, Time, and Deity* (1920), and *The Mind and its Place in Nature* (1925) respectively. I believe that these three works are

representative of the richest intellectual period of emergentism, and are sufficient for introducing the relevant themes and ideas of emergentism for our purposes here.

It is also important to note the historical context from which these works arose. Both Alexander and Broad were working during the time of a fierce debate between the theories of *mechanism* and *vitalism*. *Mechanism* states that a thing's actions can be fully explained by the action of its parts, or by some external source acting on those parts. To put it eloquently, "the whole is simply the sum of its parts." *Vitalism* stood in opposition to this idea. *Vitalism* states that organic phenomena cannot be fully explained without an appeal to some non-physical element that makes the living fundamentally different than the non-living, because these two are governed by different principles. Hans Driesch called this non-physical factor 'entelechy,' whereas Henri Bergson used the term 'élan vital.' These two were considered vitalism's most prominent defenders.

Emergentism was seen as an alternative view that offered a path between these views. It was naturalistic in the sense that it did not postulate any nonphysical entities into its metaphysics such as entelechies, élan vital, or any Cartesian Soul. Further, emergentism still maintained a hierarchical view of nature giving a unique role to living phenomena that separated it fundamentally from the mere mechanics of the micro world. To put it briefly, emergentism states that all living and non-living beings and structures are composed of the same basic elements, but emergent entities "arise" out of the basic elements and are yet "novel" and irreducible with respect to them. (Stanford Encyclopedia of Philosophy, *Emergent Properties*)

It should be noted that this section is meant to be a brief historical overview meant to familiarize us with some concepts and concerns that will come later. As mentioned

before, I will not try to present a comprehensive catalog of the entire history of emergentism. It should also be noted that there is some disagreement over some details below, but I will withdraw myself from what could be a very interesting philosophical discussion because, once again, this is not the focus of this project.

2.1 John Stuart Mill: *A System of Logic*

While Mill never used the term "emergence," his work *A System of Logic* is generally considered the platform on which the emergentist tradition was later built. This trend in thinking was most likely started by Lloyd Morgan:

> The concept of emergence was dealt with (to go no further back) by J.S. Mill in his Logic (Bk. III ch. vi 2) under the discussion of "heteropathic laws" in causation. (1923, p. 2)

The term *heteropathic laws* shows up in "Of Composition of Causes," but to understand the term *heteropathic laws*, it is useful to first present its contrast, *homopathic laws*. *Homopathic laws* follow what Mill terms the principle of *Composition of Causes*. Mill explains it as follows:

> If a body is propelled in two directions by two forces, one tending to drive it to the north and the other to the east, it is caused to move in a given time exactly as far in both directions as the two forces would separately have carried it; and is left precisely where it would have arrived if it had been acted upon first by one of the two forces, and afterward by the other. This law of nature is called, in dynamics, the principle of the Composition of Forces: and in imitation of the well-chosen expression, I shall give the name of the Composition of Causes to the principle which is exemplified in all cases in which the joint effect of several causes is identical with the sum of their separate effects. (Mill, 1843, 370-371)

The term *Composition of Causes* comes from the term "Composition of Forces" in physics, and, in turn, the paradigmatic example of the *Composition of Causes* is the law of vector addition of forces. Thus, one effect of multiple causes is identical to the sum of the multiple effects. Another way to think about it is imagine that some object has two forces pushing in different directions acting on it simultaneously. If this situation had been altered

slightly and the two forces acted on the object at two different times, and the object still ended up in the same position as in the first situation, then this is an example of what is known as a *homopathic effect*. It is an effect of multiple causes produced in the mechanical mode. Laws governing these types of effects are known as *homopathic laws*. This is the straightforward type of causation that many of the effects and laws of nature follow.

However, According to Mill, the Composition of Causes is not universal, though. His paradigmatic example is the chemical world. He states:

> The chemical combination of two substances produces, as is well known, a third substance with properties different from those of either of the two substances separately, or of both of them taken together. (Ibid. 1843, P. 371)

Consider this chemical process:

$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$

This chemical equation states that methane plus oxygen produces carbon dioxide and water. It is clear that the product is not in any sense the sum of the effects of each resultant. Mill says that this is characteristic of chemical processes, and it becomes even more vivid when we consider organic phenomena. "The phenomena of life… bear no analogy to any of the effects which would be produced by the action of the component substances considered as mere physical agents" (Ibid. p. 371). Mill calls these types of instances *heteropathic effects* and the laws that assert causal relations between causes and heteropathic effects of the causes are called *heteropathic laws*. Heteropathic laws owe their existence to a breach of the Composition of Causes and the existence of a breach of the Composition of Causes then explains the existence of the special sciences.

Homopathic effects later came to be known as *resultant* effects, and heteropathic effects later came to be known as *emergent* effects. These two terms come from George

Henry Lewes, and this is generally considered Lewes' biggest contribution to Emergentism (McLaughlin, 1992, 65). Further, the distinction between homopathic laws and heteropathic laws is generally considered as Mill's significant contribution to emergence, but many have argued his influence goes much further than this. Mill is also responsible for defining the distinction between ultimate and derivative laws. "A science is deductive in Mill's sense, if it contains a small group of systematically well-integrated laws from which all its other laws can be derived" (Ibid., 1992, 28).

Mill believed chemistry was far from being deductive, although he maintained that this was an empirical question. In fact, he was very open to the possibility of chemistry and physiology eventually becoming deductive sciences, but this can only be discovered inductively rather than deductively. Conversely, he held tightly to the idea that the laws of life would never be reducible.

> The ultimate Laws of Nature cannot possibly be less numerous than the distinguishable sensations or other feelings of our nature;- those, I mean, which are distinguishable from one another in quality, and not merely in quantity or degree. For example; since there is a phenomenon *sui generis*, called colour, which our consciousness testifies to be not a particular degree of some other phenomenon, as heat or odour or motion, but intrinsically unlike all others, it follows that there are ultimate laws of colour; that though the facts of colour may admit of explanation, they never can be explained from laws of heat or odour alone, or of motion alone, but that however far the explanation may be carried, there will always remain a law of colour. I do not mean that it might not possibly be shown that some other phenomenon, some chemical or mechanical action for example, invariably precedes, and is the cause of, every phenomena of colour. But though this, if proved, would be an important extension of our knowledge of nature, it would not explain how or why a motion, or a chemical action, can produce a sensation of colour […] (Mill, 1843, p. 485).

This should look strikingly similar to the current arguments about the irreducibility of *qualia*, which is quite remarkable considering the debate on qualia is a fairly recent one. The details of this debate will be relevant in the following sections, thus, I will not say more about it here. I just wish to add further context to the historical development of emergence.

2.2 Samuel Alexander- *Space, Time, and Deity*

Samuel Alexander's thought is distinctive within the British Emergentist tradition because of its highly metaphysical flavor. He was also the first to explicitly postulate a hierarchical view of the world, which is now a central theme in emergentism. The lowest level is space-time, from space-time, matter, from matter, life, from life, consciousness, and finally, from mind, the quality of Deity. Now, it may be useful to examine what Alexander means by *emergence*:

> The emergence of a new quality from any level of existence means that at that level (the new level) there comes into being a certain constellation or collocation of the motions belonging to the level, and this collaction possesses a new quality distinctive of the higher complex. The quality and the constellation to which it belongs are at once new and expressible without residue in terms of the processes proper to that level from which they emerge. (1920, p. 45)

From here, we can see Alexander is speaking of a new "emergent" quality. Further, Alexander states:

> The higher-quality emerges from the lower level of existence and has its roots therein, but it emerges therefrom, and it does not belong to that lower level, but constitutes its possessor a new order of existent with its special laws of behavior. The existence of emergent qualities thus described is something to be noted, as some would say, under the compulsion of brute empirical fact, or, as I should prefer to say in less harsh terms, to be accepted with the "natural piety" of the investigator. It admits no explanation.
>
> To adopt the ancient distinction of form and matter, the kind of existent from which the new quality emerges is the "matter" which assumes a certain complexity of configuration and to this pattern of universal corresponds the new emergent quality. (1920, p. 46-47)

Essentially, what Alexander is trying to express in these two passages is that all emergent qualities are *novel* and *unpredictable*. The quality is *novel* in the sense that the quality had not occurred before and it is unpredictable in the sense that the quality could not have been predicted even with all the relevant information. The quality cannot be explained any further than this and must be accepted with what Alexander calls "natural piety."

This seems straightforward enough, but he denies any primitive form of causality in addition to physics (McLaughlin, 1992, p. 32). This means that, theoretically, the Laplacian

demon (a perfectly knowledgeable being) can predict any later physical state, given all the information. This then commits Alexander to a strong determinism. Implicit in his earlier formulation of emergent qualities is the claim that not even the Laplacian demon could predict emergent qualities and processes of minded systems. These two claims seem be in deep conflict, so much so that many philosophers are uncertain as how to resolve the apparent conflict (Ibid., p. 31). I've come across a few different attempts to solve the problem, and the one that I find most promising is the addition of a supervenience claim.[1]

Within this picture, emergent qualities supervene on a distinctive kind of physico-chemical process. What this means is that the emergent qualities display their own activity, but in accordance with physics. Thus, the structure from which the emergent quality later arises could in principle be predicted before they first appear by the Laplacian demon. It still stands that, even though the new structure itself is in principle predictable, the structure *having* some emergent property is in principle unpredictable from the complete knowledge of the structure's base components, according to Alexander.[2]

2.3 C.D. Broad- *The Mind and its Place in Nature*

The work of C.D. Broad is generally considered the peak of British Emergentism, partly because much of his thought is still relevant to the current debate on emergence. Broad's book *The Mind and its Place in Nature* was based on the Tarner lectures he delivered in

---

[1] Much more needs to be said about the concept of supervenience to fully understand it, and this project will indeed be taken up in section 3.3 and elsewhere.
[2] The kind of predictability of interest here is *theoretical predictability*. Even emergent properties are *inductively predictable*. If we always observer some emergent property in conjunction with some base microstructure, then whenever we can predict the appearance of the base microstructure, we can reasonably predict the appearance of the corresponding emergent property. More will be said about this in section 3.4.

Cambridge in 1923. These lectures were focused on one general question; can we, and if so, how, unify all the sciences? This question was inspired by the large debate between mechanism and vitalism. Broad rejected both theories, and rather offered an emergent theory that was seen as a way between the two extremes. The emergence theory maintained a naturalistic approach because it did not postulate any non-physical entities, but it also gave a real causal role to life and consciousness.

Broad distinguished three different possible types of theories that could account for the characteristic differences of behavior. The first is the kind of theory that holds that "the characteristic of behaviour of a certain object or class of objects is in part dependent on the presence of a peculiar component which does not occur in anything that does not behave this way" (Broad, 1925, p. 55). An example of this kind of theory would be *substance vitalism*. Substance vitalism holds that the necessary factor for explaining a living object's behavior is an "entelechy," which would also qualify as the 'peculiar component' that occurs only in organic matter.

The other two possible theories deny the need for an appeal to some component to explain behavior. Rather, they aim to explain the differences wholly in terms of difference in structures. The first theory of this type is the theory of emergence. They state:

> The characteristic of behaviour of the whole could not, even in theory, be deduced from the most complete knowledge of the behaviour of its components, taken separately or in other combinations, and their proportions and arrangements in the whole. (Ibid., 59)

The third type of theory stands in contrast to this definition. This type of theory state that the behavior of the whole *could*, at least in theory, be *deduced* from a sufficient knowledge of the behavior of the components. These types of theories can be considered mechanistic theories. An obvious example to which this kind of theory applies is any sort of mechanical

device, such as a clock. No one doubts that with sufficient knowledge of the clock's cogs and parts, one can understand the behavior of the clock as a whole. Contemporary reductive physicalism also states that the behavior of the whole can be explained in terms of its components, so it seems it would count as this type of theory, too.

Now let's further examine Broad's definition of emergence:

> Put in abstract terms the emergent theory asserts that there are certain wholes, composed (say) of constituents A, B, and C in a relation R to each other; that all wholes composed of constituents of the same kind as A, B, and C in relations of the same kind as R have certain characteristic properties; that A, B, and C are capable of occurring in other kinds of complex where the relation is not the same kind as R; and that the characteristic properties of the whole R(A, B, C) cannot, even in theory, be deduced from the most complete knowledge of the properties A, B, and C in isolation or in other wholes which are not of the form R(A, B, C). (Ibid., P. 61)

Thus, a property is emergent if and only if it cannot be deduced from the most complete knowledge of its properties and its components in isolation or in other wholes. One immediate question that may arise from this definition is: why the need for the expression "in isolation or in other wholes?" This is simply a way to limit what properties can be used in an attempted deduction. Otherwise, it would seem very unlikely that there could ever actually be an emergent property. The idea of the argument is that one of the kinds of properties that an object can have is that, under certain circumstances, it becomes a part of a compound with some other certain properties. Thus, this includes all "potential properties."

According to Broad, a law that connects an emergent property of a structure with the properties of the components of the structure is a *unique*, *ultimate*, and *irreducible* law. It is not simply some special case of a more general law, and, further, it does not arise from a combination of more general laws. It's a law that can only be discovered through the study of the particular case (Ibid., p. 65). This then essentially corresponds to what Mill had in mind with his "ultimate laws."

Now, it may be helpful to turn to Broad's mathematical archangel example that illustrates some of these concepts. This example came from his discussion on the theoretical limitations of prediction. The mathematical archangel is intended to be a being with perfect mathematical skills and also possesses all the relevant information about the universe. According to Broad, if the emergent theory of chemical compounds is true:

> A mathematical archangel, gifted with the further power of perceiving the microscopic structure of atoms as easily as we can perceive hay-stacks, could no more predict the behaviour of silver and chlorine or the properties of silver-chloride without having observed samples of those substances than we can at present (Ibid, p. 71).

Further, even if the mechanistic theory of chemistry turns out to be true, there is still an additional theoretical limitation to the deduction of properties of chemical elements and compounds:

> Take any ordinary statement, such as we find in chemistry books; e.g., "Nitrogen and Hydrogen combine when an electric discharge is passed through a mixture of the two. The resulting compound contains three atoms of Hydrogen to one of Nitrogen; it is a gas readily soluble in water, and possessed of a pungent and characteristic smell." If the mechanistic theory be true the archangel could deduce from his knowledge of the microscopic structure of atoms all these facts but the last. He would know exactly what the microscopic structure of ammonia must be; but he would be totally unable to predict that a substance with this structure must smell as ammonia does when it gets into the human nose. The utmost that he could predict on this subject would be that certain changes would take place in the mucous membrane, the olfactory nerves and so on. But he could not possibly know that these changes would be accompanied by the appearance of a smell in general or of the peculiar smell of ammonia in particular, unless someone told him so or he had smelled it for himself. If the existence of the so-called "secondary qualities," or the fact of their appearance, depends on the microscopic movements and arrangements of material particles which do not have these qualities themselves, then the laws of this dependence are certainly of the emergent type (Ibid, P. 71-72).

This very may well be the first formulation of the now-famous "knowledge argument." Both Thomas Nagel (1974) and Frank Jackson (1982, 1986) have championed a view very similar to this, albeit there are some very important differences. The point of the knowledge argument is to show that, even a person with complete knowledge of all the physical sciences would not necessarily know everything there is to know. To draw from Jackson's arguments, even a person with perfect knowledge of the science of optics and colors would not know what it is like to see the color red had they been locked in a black and white room

for their entire life. The conclusion of the argument is then that physical sciences are not sufficient for knowing the whole truth of the world.

Broad, like Mill, also believed that the possibility of chemistry and biology being deductive sciences was an empirical question. In fact, like Mill once again, Broad was quite open the idea of these sciences eventually becoming reductive. He still maintained that they could not be whole truths of the material world, though. Broad believed that "secondary qualities" could never be explained mechanistically. Some examples of secondary qualities would be smell, tastes, colors or essentially any other sense experience. The laws connecting microscopic particles and secondary qualities must be emergent laws, and a complete account of the world cannot ignore these laws, according to Broad (Ibid, p. 72).

Broad thought that mechanism would introduce a deep unity into the external world and, consequently, the sciences that deal with it. According to emergentist theories, on the other hand, the external world and the sciences would form a type of hierarchy. This should sound very similar to the ideas of Alexander, although they each gave different labels to each level of the hierarchy. Like Alexander, Broad's hierarchy is still consistent with the idea that there is only one fundamental kind of stuff that composes the world (Ibid., p. 77). Even with this principle holding, we must recognize that, within the stuff, there are aggregates of different orders. Consequently, this then requires two kinds of laws.

Broad called these two types of laws *intra-ordinal laws* and *trans-ordinal laws*. *Intra-ordinal laws* relate the properties of the same order. These could then be considered the laws of the special sciences. He never speaks of these laws themselves as emergent, thus he most likely did not think special science laws are emergent (McLaughlin, 1992, p. 42). He believed

they would be deducible from lower-level laws and conditions together with trans-ordinal laws.

Trans-ordinal laws are the kinds of laws that connect the properties of aggregates of adjacent order (Broad, 1925, p. 78). It seems as though he thought all trans-ordinal laws were also emergent laws. According to Broad, "a trans-ordinal law would be a statement of the irreducible fact that an aggregate composed of aggregates of the next lower order in such and such proportions and arrangements has such and such characteristic and non-deducible properties" (Ibid., p. 78). This then seems to be, by definition, an emergent law. At the very least, all the laws that connect properties of the aggregate of the next lower level with secondary qualities are *necessarily* emergent. Once again, this is very similar to what Mill had in mind with his "qualia-style" argument above.

Finally, for Broad, there are three different kinds of properties: *reducible*, *ordinally neutral*, and *ultimate*. *Reducible properties* are properties that are characteristic of a certain order, but in theory could be deduced from the structure of the whole and the properties of its constituents. *Ordinally neutral properties* are properties that are shared across orders. Examples of what kinds of properties fit this definition according to Broad are inertial and gravitational mass (Ibid., p. 79). Further, physics concerns itself with ordinally neutral properties, or the most general characteristics of matter. Finally, *ultimate properties* are properties of a certain order that all aggregates of this order, but no aggregates of lower orders possess, and which could not be deduced from the structure of the aggregate and the properties of its constituents (Ibid., p. 78). Ultimate properties are then clearly emergent properties for Broad.

2.4 The Fall of British Emergentism

During the 1920's, British Emergentism was widely discussed in the mainstream of philosophy. Emergence itself was the theme of several symposiums and the British Emergentists themselves acquired a mild fame. Contrarily, emergentism was also widely criticized, but it was not philosophical criticism that pushed emergentism out of philosophical favor. In fact, "no fundamental inconsistency or philosophical error was ever pointed out" (McLaughlin, 2008, p. 92). This may lead one to wonder, "what led to the fall of British Emergentism if it wasn't philosophical error?" I believe the answer to this question is two-fold.

Brian McLaughlin (2008) points out advances in science as the main factor for the downfall of the British tradition. Shortly after Broad's seminal work *The Mind And its Place in Nature* was published, quantum mechanics was discovered. Quantum mechanics made possible, among other things, the explanation of chemical bonding in terms of electro-magnetism. This then catalyzed the development of molecular biology and, eventually, the discovery of DNA. In effect, all of these discoveries reduced both biological and chemical properties to further base components. The obvious result of this is that properties that fall within these levels are then very likely *not* truly emergent properties. The British Emergentists championed properties in the biological and chemical realm as paradigmatic examples of emergent properties. Given that their favorite empirical examples of emergent properties were later reduced to further base components, its fairly clear why British Emergentism lost favor within philosophical circles.

Another contributing factor to the fall of British Emergentism was the rise of logical positivism as a predominant philosophical style. From the 1930's to the 1960's, the Anglo-

American philosophy of science was highly influenced by the positivist and hyper-empiricist views of science. The anti-metaphysical consequence of such a view resulted in emergence being viewed as a vague philosophical concept that held no place in any true view of science (Kim 2008, p. 127).

2.5 The Re-Emergence of Emergence

I believe these two factors in conjunction explain why the British Emergentist tradition dwindled from the period right after its heyday (1920's) up until about the 1960's. During this period, emergentism did not die out; rather, it seems that it merely lost its place within the center of academic philosophy. As noted earlier, the chief examples of emergent properties put forth by the British Emergentists (biological and chemical properties) now seem very unlikely to be truly emergent properties due to scientific advances, but none of *these* advances in science dangerously affect the plausibility of emergent properties in the mental realm. It was right around the 1960's when the concept of emergence started to find its place in philosophy of mind. Failure of reductive explanation within the discipline led to the rise of "non-reductive physicalism," which has been argued by many, and possibly most famously by Jaegwon Kim, to simply equate to emergentism. If this assertion is true, then emergentism has been central in discussions of philosophy of mind since the early 1970's (Kim 2008, p. 128).

3. The Central Characteristics of Emergence

The aim of this chapter is to critically examine some of the central themes within theories of emergentism. The ultimate goal of this chapter is to build a comprehensive definition of an emergent property from the discussion of the most important characteristics, and each section within this chapter will add one requirement to the definition of an emergent property. By the end of the chapter, we will have a complete definition. Part of the reason why emergence is thought of as a vague concept is because there are many different types of emergence varying across the philosophical landscape. Some forms of emergence are fairly radical while others are rather mundane. This being said, I will not be after just any definition of an emergent property; rather, I will look to create a definition that has the most significance to philosophy of mind. Further, I will be after a metaphysically interesting theory of emergence, which will involve some kind of ontological claim that validates a certain autonomous existence of emergent phenomena. To accomplish such a task, I will follow closely the accounts of David Chalmers, Jaegwon Kim, and Brian McLaughlin on five different characteristics of theories of emergence: levels of complexity, genuine novelty, supervenience, irreducibility/unpredictability, and downward causation.

3.1 Levels of Complexity

One of the very basic ideas of emergence is that, there exists in this world, different levels of complexity. The emergentist claims that, throughout the history of evolution, there came into being different structures that increased in complexity, thus creating a hierarchy. This can be seen in the following passage from Broad: "we have to reconcile ourselves to

much less unity in the external world and much less intimate connection between the various sciences. At best the external world and the various sciences that deal with it will form a hierarchy" (Broad, 1925, p. 77). Broad is suggesting here that the special sciences have an autonomous existence in the sense that the purpose of some special science is not to simply create concepts and a language to help us understand seemingly different levels of complexity while the special science in question ultimately has to answer back to the fundamental science of physics.

Rather, each special science deals with its own level of distinct properties. The level of existence to which something in nature belongs is determined by its characteristic property. An example of such a characteristic could be vitality, and a property that has this characteristic would belong to the level of the biological, or mentality, and a property that has this characteristic would belong to the level of the psychological. These characteristic properties are wholly made up of kinds of lower orders, but yet are specific properties of a given order. Broad called these *ultimate characteristics* of the order. Entities at the very basic level have no structure themselves, but are the building blocks for all structures that follow. There is then a fundamental part-whole relation within the hierarchical model.

The most fundamental level is always that of the physical, and, once again, everything that comes after will always be made up completely of elements from this level. This then commits all emergentists to *physical monism*, which states that all entities in the world are composed of physical elements. This kind of doctrine leaves no room for any appeal to anything supernatural; "there are, for example, no Cartesian souls, or entelechies, vital élan, or the like" (McLaughlin, 1992, pg. 19). The benefit of this doctrine is it allows for an empirical, scientific viewpoint.

Each emergentist has his or her favorite hierarchy of levels of complexity. For Broad, it started with physics, then chemistry, then biology, and finally psychology. Alexander's went from space-time, then matter, then life, then consciousness, and, finally, the quality of deity. As highlighted in the previous chapter, we now know that it is very unlikely that either biological or chemical properties are truly emergent, but, if the British emergentists were right about anything, they were right to point out properties of consciousness as candidates for a truly emergent property (Kim, 2008, p. 131). This is also our interest here, so we only have to assume there are *at least* two levels: the level of the physical properties and the level of conscious properties. I say *at least* because, if there really are different levels of existence, these two are sure to be part of it and the discovery of additional levels is no danger to the emergentist's claim that conscious properties are truly emergent.

From here, we know that the emergentist holds a layered view of nature, and they are also committed to physical monism. Because the emergentist is committed to physical monism, we know that each level is made up of elements from the most basic level (i.e. the physical). From here, we can add our first requirement to our working definition of an emergent property. The rest of the requirements will be added through the coming sections in this chapter. P is an emergent property if and only if i) *P is composed of elements completely from lower levels and...*

3.2 Genuine Novelty

Another central tenet to many theories of emergence is the idea that what emerges is something *genuinely novel*. This notion is inherently vague; what is meant by something novel

and what makes it genuine? Does it refer to new structures, new entities, or new properties? Could it even mean something like new laws?

One way to go about answering this question is to establish what is not meant by the emergentist when they claim what emerges is genuinely novel. Emergent properties are unlike those that are merely *resultant*. The properties of resultants can be known from a sufficient knowledge of the base components. This then rules out something like "sphere-ness" as an emergent property when instantiated for the first time, precisely because the property of "sphere-ness" can be known from the knowledge of the object's base properties. For the same reason something like "sphere-ness" would be considered a resultant property and not an emergent one, numerically new identity is also ruled out as a genuinely novel property. For example, the first table that weighed 35.2981475 lbs is not new in the sense invoked by the emergentist. What we're after is a new *type* that has never been instantiated before. This can mean the instantiation of new type of *structure* or new type of *property*.

Jaegwon Kim suggests the term "novel" has two dimensions when used by the emergentist. First, there is the epistemological sense. This sense of the term "new," means that something is new because it is unpredictable. Further, there is also a metaphysical sense, which states that an emergent property brings with it new causal powers, or powers that did not exist before its emergence (Kim, 2008, p. 131). Clearly, the latter is a much stronger formulation of an emergent property. Moving forward from here, our project is to examine which notion of "novelty" is most significant to philosophy of mind. The next three sections on *logical* and *nomological supervenience*, *unpredictability/irreducibility*, and *downward causation* in combination will give us a clear answer to this question. All we really need here is a rough notion of novelty; novelty refers to either new *types* of structures, or more importantly here,

new *types* of properties. All that we can add to our working definition at this point is a self-evident and fairly non-illuminating requirement; P is an emergent property if and only if i) P is composed of elements completely from the lower level, ii) *P is a genuinely novel property and...* The next three sections will add more detail to this claim.

3.3 Logical and Nomological Supervenience

While the British Emergentists never used the term *supervenience*, it is a central concept in contemporary discussion of emergence. It seems fairly clear that all emergentists past and present were/are committed to at least some form of supervenience, though (Kim, 2008, p. 130). Holding a supervenience relation is a way to hold that mental properties are *distinct* from the physical, but are nonetheless *determined* by the physical. A supervenience thesis is then the marriage of two other central tenets of the British Emergentists: *determinism*[3] and *unpredictability.*[4] To put it in our context, mental properties are unpredictable on the basis of physical properties because they are distinct from them, but mental properties are dependent on the physical because a mental change cannot occur without a physical change, thus there is a dependence relationship of the mental on the physical.

In general, supervenience is a relation between two-sets of properties: *B-properties* and *A-properties*. Intuitively, the B-properties correspond to *high-level* properties and A-properties correspond to the most basic *low-level* properties (Chalmers 2006, p. 33). In the context of

---

[3] This is not the type of determinism that is associated with debates of free will. That type of determinism is of the ontological sort, whereas this sort of determinism is meant to highlight a dependence relation.

[4] Supervenience itself does not guarantee unpredictability. Certain supervenient properties can be predicted from full knowledge of its base property, but there is a certain class of supervenient properties that cannot be predicted from knowledge of its respective base properties. Thus, supervenience can allow unpredictability. More will be said about the distinction between these two types of supervenient properties later in this section.

philosophy of mind, A-properties can be thought of as *physical* properties, and B-properties can be thought of as *mental* properties. Our general template for the definition of supervenience will go as follows: B-properties supervene on A-properties, if and only if no two possible situations are identical with respect to their A-properties while differing in their B-properties (Ibid., p. 34).

There are two types of supervenience that will be of interest to us here. The first is *logical supervenience* and the second is *natural* (or nomological) *supervenience*. B-properties supervene *logically* (or conceptually) on A-properties if no two *logically possible* situations are identical with respect to their A-properties but are distinct with respect to their B-properties. It is important to note here that logical supervenience is not defined in terms of deducibility in any system of formal logic. Rather, logical supervenience is defined in terms of logically possible *worlds*. The test for logical supervenience is to ask if the set of A-facts in question conceptually *entail* the B-facts in question. If the A-facts do indeed entail the B-facts, then we can say the B-facts supervene logically on the A-facts. In this sense, entailment is necessary and sufficient for logical supervenience. Another way to think about it is, if B-properties supervene logically on A-properties, then once God creates a world with certain A-facts, the B-facts come along as an automatic consequence (Chalmers, 2006, p. 35).

In general, B-properties supervene *naturally* on A-properties if any two *naturally possible* situations with the same A-properties have the same B-properties. A naturally possible situation is one that could actually occur without deviating from the natural laws of our world. This then corresponds to what we think of as real empirical possibility. Or, in other words, it can be thought of as a naturally possible situation that could actually occur in the real world, given the right conditions. Naturally possibility is sometimes referred to as

nomic or *nomological* possibility. Conversely to our example above, if B-properties merely supervene naturally on A-facts, then once God created a world with certain A-facts, God then had to make sure there was a law relating the A-facts and B-facts (Ibid., p. 34).

There are nearly a countless number of situations that are logically, but not naturally, possible. Contrarily, any situation that is naturally possible will also be logically possible. Thus, naturally possibility is a sub-class of logical possibility. Natural supervenience holds when, among all naturally possible situations, those with the same A-properties have the same instantiation of B-properties. Thus, natural supervenience is in place when the B-facts about a situation are *naturally necessitated* by the A-facts. Natural necessitation occurs when, among all naturally possible situations, the same group of A-properties is always in conjunction with the same B-properties. Further, the connection between the two is lawful-not merely coincidental, in the sense that when the A-properties are instantiated, they will always bring about the B-properties (Ibid., p. 37).

It is rather difficult to find cases of natural supervenience on some set of physical properties without logical supervenience, but consciousness itself may be the best example:

> "It seems very likely that consciousness is naturally supervenient on physical properties, locally or globally, insofar as in the natural world, any two physically identical creatures will have qualitatively identical experience. It is not at all clear that consciousness is logically supervenient on physical properties, however" (Ibid., p. 37).

From this point here, Chalmers constructed a very popular and forceful thought experiment involving philosophical zombies. In some possible world, if it is possible that there can be a creature physically identical to a conscious creature but have no conscious experience, then, as a consequence, conscious experience does not supervene logically on the physical. This then means that the correlation between physical structure and conscious

experience is not ensured by any logical or conceptual means. Rather, our laws of nature ensure that connection (if conscious experience truly does supervene on the physical).

From this discussion of logical and natural supervenience above, it should start to be fairly clear how supervenience relates to genuine novelty. It will hopefully be clear after section 3.4. For now, we should be able to see that, if some property P logically supervenes on some microstructure M, then P is merely a resultant property, which we should be familiar with from the previous chapter. If some property P merely naturally supervenes on some microstructure M, then P is still a candidate for being an emergent property, and thus a genuinely novel property. A lack of logical supervenience is then *necessary* for, but not *sufficient,* for some property's being genuinely novel. The rest of the necessary components will come from the three following sections on *irreducibility*, *unpredictability*, and *downward causation*. Now, our definition reads P is an emergent property if and only if i) P is composed of elements completely from the lower level, ii) *P supervenes on M with nomological necessity but not logical necessity and...*

3.4 Unpredictability/Irreducibility

While these two main themes are generally considered distinct concepts within the emergentist tradition, I find it useful to treat *irreducibility* and *unpredictability* as forming a single package. Both concepts have potentially different implications, but for what is trying to be argued here, the two have a very close tie. One of the main claims of the emergentist is that emergent properties are not *reductively explainable* in terms of the basal constituents from which they arise. This claim is often followed by the further claim that emergent properties

are *not predictable* from their basal conditions, even with an exhaustive knowledge of the base conditions.

Unpredictability was a main concept for Alexander's more diachronic theory of emergence, but current discussion in philosophy of mind is more interested in the synchronic theories like the one put forth by Broad discussed above. In those kinds of theories, emergent properties are unpredictable *precisely because* they are irreducible.[5] This then means that irreducibility is the most significant form of unpredictability for current theories of emergence in philosophy of mind.

There are a couple of ways that something can be reducible yet unpredictable: for example, extreme chaotic systems (Braddon-Mitchell, 2007, p. 289). The calculation involved in predicting such a system is hypersensitive to the accuracy of measurement of initial conditions, such that, a small error in initial measurement will result in a massive difference in the macroscopic properties that arise. Given that there is no way in principle to get zero error, these kinds of properties are unpredictable in principle even if they are reductive in principle. While this is an intriguing result, it does not seem to bear any theoretical significance in relation to the prospects of truly emergent ontologically mental phenomena.

A better place to start may be by distinguishing between two types of predictability: *inductive predictability* and *theoretical predictability*. Even emergent properties are inductively predictable. If we always observe some emergent property *E* in conjunction with some property *M* in some system, then any time we either know or believe that system will

---

[5] This very well could be the case for theories of emergence like Alexander's as well, but he never emphasized this point.

instantiate that same microstructural property *M* at some time *t*, then we can predict that that system will also instantiate emergent property *E* at time *t*. To put this more generally, from empirical data we may have a well-developed "emergent law" that connects certain microstructural properties with emergent properties. What the emergentist is denying is the theoretical predictability of *E* on the basis of *M* (Kim 2008, p. 131).

This can be seen by their invocation of such things as archangels and Laplacian demons, which both had unlimited factual information as well as unlimited cognitive resources. Despite their complete knowledge, even beings with these capabilities cannot predict the appearance of emergent properties before they first appear. The explanation for this peculiar result is that emergent properties are irreducible properties, in the sense that they are not reductively explainable in terms of their basal constituents, and things that are not reductively explainable in terms of their basal constituents are also theoretically unpredictable.

"In general, a reductive explanation of a phenomenon is accompanied by some rough-and-ready analysis of the phenomenon in question, whether implicit or explicit" (Chalmers, 1996, p. 44). To align an example with our early discussion, let us consider an example borrowed from Jaegwon Kim- the reduction of a gene to the DNA molecule. The notion of being a gene can roughly be analyzed in terms of having some property that performs the causal function of transmitting phenotypic characteristics from parents to offspring. This is the conceptual component of our analysis. It follows that once we've explained the process by which phenotypic characteristics are transmitted from parents to offspring, we have then reductively explained that instance. This is the empirical component of our analysis. It just so happens that the DNA molecule fills the causal role highlighted

above, and we have a fully developed theory that explains how the DNA molecule performs this causal work. Thus, we are allowed to claim that the gene has been reduced to the DNA molecule (Kim, 2008, p. 132).

The type of explanation above is a *functional analysis*. The core of such notions can be characterized in terms of the performance of some function or functions, or in terms of the capacity to perform those functions. For example, functional analysis shows that the disposition of solubility of water is the state of being in a state that disposes its occupants to dissolve in water (McLaughlin, 2008, p. 89). Once we've explained how some thing performs some function, then it follows that the phenomena in question has been explained. In this sense, there is no need to *identify* some B-fact with some lower-level A-fact, thus, *multiple realizability* itself does not rule out the possibility of a reductive explanation via functional analysis. As long as the results of our causal story are physically characterizable, then there should be a physical explanation for the performance of that type of function. The empirical component of the analysis will always constitute the vast majority of the reductive explanation, whereas the conceptual component will be largely trivial. Once the necessary empirical data are in place, we can then create a story about the lower-level physical causation. Once we have explained how the relevant causal work is performed, the phenomenon of interest can be reductively explained.

Many mental concepts can be analyzed functionally just like the human gene. To explain these kinds of states, we give a story about how the underlying components perform the relevant causation. In this case, we give this story by explaining the underlying neurophysiology of some mental concept. Let us take the concept of learning, for example. We have at our disposal an account of the conceptual component (it is the acquirement or

modification of knowledge, behaviors, skills, etc.) and the underlying neurophysiology of learning (it is the change in structure and action of neurons such that they hold information in long term memory in the temporal and parietal lobes of the cortex). With the two components together, we have the ability to give a physical account of learning by creating a functional model.

When we consider *phenomenal states*, the kind of causal story that works well in explaining psychological states does not seem to apply very well. Whatever functional account of human cognition we can give, there is still a *further question*; why is this kind of functioning accompanied by consciousness (Chalmers, 1996, p. 47)? This kind of question does not arise when we are analyzing psychological states. Let us consider the functional model of learning again. If we ask the question, "why is that sort of functioning accompanied by learning?" we are left with a purely semantic answer. All it *means* to learn is to function in that sort of way. There is no corresponding analysis of the concept of consciousness, because phenomenal states are not defined by their causal role. Thus, there is what is referred to as an *explanatory gap*. Even if we create a functional model that is able to give rise to consciousness, the question of *why* it gives rise to consciousness remains unanswered because we are ignorant of both the conceptual and empirical component of consciousness (Ibid., p. 48).

Strictly, speaking only laws, not properties themselves, can be reduced. This is because, when we talk about supervenience relations, we are implying supervenience principles or laws that state if the system has some set of A-properties, then the system will also have some corresponding set of B-properties. These are what Broad referred to as *trans-ordinal laws*. The key issue here is whether these supervenience laws relating physical facts

with phenomenal consciousness are fundamental, irreducible laws or whether they are actually just derivative some lower level law. "A law L is fundamental if and only if, it is not metaphysically necessitated by any other laws, even together with initial conditions" (McLaughlin 2008, p. 93). It should be clear that, if the supervenience law relating physical facts with phenomenal consciousness is a fundamental, irreducible law, then phenomenal consciousness is a truly emergent property. By definition, if something is a truly emergent property, then it is also irreducible to some set of lower-level facts.

Now we can see how closely supervenience and irreducibility are related in the relevant discussion here, so it may be helpful to put some of these issues surrounding irreducibility in the context of our earlier discussion of supervenience. When we instantiate some appropriate functional model of learning, it is logically impossible that account will be instantiated without learning. Philosophers like Chalmers will argue that it does seem that the account of learning could be instantiated without any accompanying phenomenal consciousness, though. If some phenomenon is logically supervenient on some lower-level properties, then that phenomenon seems likely to be reductively explainable in terms of those low-level properties. The reason why the human gene is reductively explainable in terms of lower-level facts is precisely because it is logically supervenient on those facts (Chalmers, 1996, p. 48).

From the arguments above, it isn't exactly clear whether logical supervenience is *sufficient* for reductive explanation, but this is only a peripheral issue. This is because all that is needed here is for logical supervenience to be *necessary*, rather than sufficient, for reductive explanation. If it is possible to create a functional model of learning that does not have any accompanying consciousness, then any proposed functional or physical account of

phenomenal consciousness will be incomplete. This is because learning without consciousness suggests that consciousness itself does not logically supervene on the lower level facts that fixed the instantiation of learning. What is required for there to be some sort of accompanying consciousness then is an appeal to some further fundamental law that will then fix consciousness to our functional model of learning.

From here we can add two more components to our working definition that further illuminate what is means for some property to be genuinely novel. P is an emergent property if and only if i) P is composed of elements completely from the lower level, and ii) P supervenes on M with nomological necessity but not logical necessity, iii) *P is not reductively explainable in terms of its basal constituents, and iv) the laws relating P to its basal constituents are fundamental, irreducible laws and...*

3.5 Downward Causation

The last characteristic of emergentism that will be discussed here is the doctrine of *downward causation*. The reason why it is discussed last is because it is the most problematic characteristic of emergence. Downward causation is a metaphysical issue, and if it proves to be an absurdity, it is argued that it will greatly damage the philosophical coherence of emergence- so much so that, if we cannot make sense of downward causation, the doctrine of emergence may very well be dealt a fatal blow. Here, I only wish to make some precursory remarks, as downward causation will be dealt with at greater length in section 6.

The irreducibility of emergent properties as formulated above seems to imply something like downward causation, or at the very least, same-level causation at the non-basic level. If the behavior of the components of the system cannot be deduced from their

behavior in other systems, something must have additional causal influence on them- most plausibly the emergent properties themselves. I then argue emergent properties having their own causal powers (not powers also possessed by its supervenience base) is then a necessary claim for any emergentist who wishes to hold an ontologically interesting theory of emergence. If they do not insist on the causal efficacy of emergent properties, then that view dissolves into *epiphenomenalism*, which is not a good result for the emergentist. Samuel Alexander was well aware of this: "(epiphenomenalism) supposes something to exist in nature which has nothing to do, no purpose to serve, a species of *noblesse* which depends on the work of its inferiors, but is kept for show and might as well, and undoubtedly would in time be abolished" (Alexander, 1927, p. 8). If we hold that mental properties are emergent properties, rather than resultant properties, but accept epiphenomenalism, then any interesting theoretical implications of emergent properties having a real causal effect on the world seem to fall away. Some philosophers (Jackson, 1982) are willing to accept epiphenomenalism of mental properties, but this account of the mental is not acceptable here. What I aim to analyze is a causal theory of emergent mental properties.

Thus, the causal efficacy of emergent properties is of great importance to the kind of emergentism that is being presented in this paper. We can formulate the idea of downward causation as follows: emergent properties are to have their own distinctive causal powers and they are also able to exercise their causal powers "downward" with respect to the lower-levels from which they emerge (Kim, 2008, p. 140). When we consider the hierarchy of complexity as formulated in the first section of this chapter, the very idea of downward causation involves vertical directionality. Some property can be located at a "higher," "lower," or "same" position within the hierarchy in relation to some other property on the hierarchy (Ibid., p. 141).

This then implies three different types of inter- or intra- level causation: i) *same-level causation*, ii) *downward causation* and iii) *upward causation*. As the name suggests, same-level causation involves a causal relation between two properties at the same level of the hierarchy. This also includes cases of when the instantiation of some emergent property causes the instantiation of some other emergent property. Upward causation causes the instantiation of some higher-level property by means of some lower-level property, and downward causation causes the instantiation of some lower level property by means of some higher-level property (Ibid., p. 142).

The concepts of upward causation and same-level causation (at the basic level, at least) are fairly unproblematic, but, as mentioned before, the same cannot be said for downward causation. The emergentist wants to claim that emergent properties have causal efficacy on lower-level components of the system (or at the very least, same-level causation by the mental), but the problem is that the laws of physics don't seem to leave any room for additional causal powers. In other words, the physical realm is *causally closed*. Jaegwon Kim best formulates this and other problems of downward causation in what is known as "the exclusion argument" (Kim, 1993). The problem goes as follows: i) if the emergentist wants to hold on to the doctrine of causal efficacy of emergent properties, then he or she must accept a form of downward causation and consequently deny the causal closure of physics, ii) if the emergentist wants to hold onto the doctrine of causal efficacy of emergent properties while also honoring the causal closure of physics, then he or she must accept causal overdetermination- an apparent absurdity, or iii) if the emergentist denies downward causation, then he or she must accept that emergent properties have no causal relevance at all. Thus, emergent properties are epiphenomenal.

It should be noted here that, although mental causation and downward causation are closely related concepts, they come apart in some very important ways. There can be mental causation without downward causation (in reductive physicalism, for instance), and downward causation without mental causation (presumably some causal influence from some higher-level factor that is not mental). It is often thought that mental causation is a form of downward causation, but this is really only true of non-reductive physicalist theories of mind, and even that need not be the case. The main concern in this paper is mental causation, but it will require some related discussion on downward causation.

This problem will be examined at great length in section 5, and challenged in chapter 6, so all I wish to do here is introduce it. For now, all that needs to be clear is that the emergentist is committed to some form of causal efficacy of emergent properties if he or she wants to hold some sort of theoretically interesting thesis. Thus we can add our final component to our definition of an emergent property: P is an emergent property if and only if i) P is composed of elements completely from the lower level, and ii) P supervenes on M with nomological necessity but not logical necessity, iii) P is not reductively explainable in terms of its basal constituents, and iv) the laws relating P to its basal constituents are fundamental, irreducible laws, and *v) P has causal powers of its own and these powers are irreducible to the causal powers of its basal constituents.*

4. Emergentism and its Relation to Physicalism

Emergentism was born in the midst of the debate between the two extremes of mechanism and vitalism. As the metaphorical offspring of these two theories, the original emergentists intended not only for their theory to inhabit a middle ground between its two ancestors, but also synthesize what is desirable from both and leave what is unfavorable out of its formulation. In this sense, the emergentist goal has both a *positive* and *negative* element to it. The negative element is a conjunction of two negative theses: don't fall into reductionism and don't posit any supernatural entities. The positive element is to save a special ontological spot for the mental while staying scientifically respectable. These commitments then essentially equate emergentism to nonreductive physicalism.[6] This is because the two theories both deny a type-type reduction of the mental to the physical. It has also been suggested that emergentism was the first systematic formulation of nonreductive physicalism. Nonreductive physicalism is now the most influential position on the relation between the physical and the mental (Kim, 2006a, p. 291). The nonreductive physicalist denies the physical reducibility of the mental, but accepts a robust and intimate relationship between mental and physical properties. They explain this connection by appealing to supervenience, which should now be a familiar concept from earlier discussion.

---

[6] The only place the two may come apart is how we interpret the type of reduction is involved with their respective irreducibility claims. We know from the arguments presented in this paper that the model of reduction most appropriate for Emergentism is the functional model of reductionism. If we apply functional reduction to another non-reductive physicalist theory such as functionalism, it turns into something much closer to reductive physicalism as opposed to non-reductive physicalism. For this reason, I contend that the irreducibility claim of nonreductive physicalist theories is best thought of as not type-reducible.

There are two main questions of interest in this section. First, is the theory of emergence that I present here justifiably a form of physicalism, or is it something closer to dualism? The second is whether or not the emergentist was successful in his or her original goal of establishing a stable position between the two extremes of mechanism and vitalism.[7] These are no doubt related questions, but they will come apart from each other in an important way.

To answer these questions, more discourse on supervenience is needed. Many think that the emergentist's commitment to supervenience legitimizes the position in the eyes of the physicalist, and it removes some of the "spookiness" stigma associated with dualism. One way to develop a further understanding on how supervenience explains the relationship between mind and body is to contrast it with a couple of other purported types of physicalism, and see which, if any, provide a theoretical grounding of how the mind relates to the body. Both *token physicalism* and *physical realizationism* will be of importance here. After further detailing supervenience as well as these two other concepts, it will be clear what theoretical implications the emergentist is committed to concerning how the mind relates to the body, and the result of this will provide sufficient grounds for answering our two questions of interest.

4.1 Token Physicalism

Token physicalism is a good place to start our discussion, because it represents an intuitive requirement for any physicalist theory.

---

[7] To put this in a contemporary context, mechanism then coincides with physicalism and vitalism coincides with dualism- specifically substance dualism.

Token physicalism- every event that has a mental property also has a physical property.

Much of our discussion has been concerned with the talk of properties rather than events, so it may be helpful to establish how *events* and *properties* are related. There are a couple of possible approaches to events, and each has its own metaphysical implications, but only the conception presented here will be necessary for our purposes. Our approach takes events as basic concrete particulars of the world, along with material objects (Kim, 2006a, p. 101). Given this definition, events, like material things, have properties. Consider a particular occurrence of pain. It is an event that falls under the event kind of pain; further, we may say this occurrence has the property of being a pain. According to token physicalism, this event also has a physical (or a neurophysiological) property: the property of C-fiber excitation. This pain event E then has both the property of being a pain and the property of being a C-fiber excitation event.

From our definition formed in section three, requirement *i* is a shared commitment of both the emergentist and the token physicalist. Both requirement *i* and token physicalism assert a kind of physical monism, thus both are incompatible with something like Cartesian substance dualism. While this may seem a good way to capture physicalism, it is actually a very weak form of physicalism. Most notably, it is consistent with certain forms of dualism- specifically, property dualism. Given this fact alone, token physicalism cannot be enough to guarantee physicalism, because we know physicalism is supposed to be inconsistent with any form of dualism. Further, token physicalism says nothing about the relationship between the mental and the physical. The theory can be true even if there is no systematic, much less law-like, connection between mental and physical properties, and the former need not depend on the latter in any way. All token physicalism tells us is that mental properties and

physical properties are instantiated by the same entities and that every one of these entities has at least one physical property, but this is clearly not enough to constitute a robust physicalist theory (Ibid., p. 104).

4.2 Supervenience as Minimal Physicalism

Thus, to have a fully developed physicalist theory, it seems as though something needs to be said about how the mental and the physical are related. Some philosophers think that supervenience is a way to explain the connection between the mental and the physical, so further discussion of the concept will be appropriate.

Supervenience physicalism- if the physical nature of a thing is fixed, so too is its mental nature (Stanford Encyclopedia of Philosophy, *Physicalism*).

As mentioned prior, there are many different forms of supervenience, but we can focus on what is known as *strong supervenience*:

> Mental properties *supervene strongly* on physical properties, in that necessarily, for any mental event M, if anything has M at time t, there exists a physical base (or subvenient) property P such that it has P at t, and necessarily anything that has P at a time has M at that time (Kim, 2000, p.9).

To state this definition more plainly, all mental properties have a physical base, and further, that physical base guarantees the instantiation of the corresponding mental property. The connection between certain mental properties and certain physical properties is one of necessity, such that, consequently any two things in the same or different possible worlds that are exact physical duplicates will be exact psychological duplicates as well.

If the mental properties of some entity are fixed by that entity's physical properties, the mental *depends* on the physical. Starting with Donald Davidson (1970), it has been customary to associate supervenience with dependence or a determination relationship, precisely for the reason stated above. There is one place where this exact association comes apart, though. A determination relation is asymmetric in the sense that if x depends on or is determined by y, it cannot be that y also depends on or is determined by x. However, mind-body supervenience as we have formulated it is not asymmetric. If x supervenes on y, this does not exclude the possibility that y in turn supervenes on x (Kim, 2000, p. 11). Despite this fact, supervenience should still be thought of as a dependence relation for our purposes here, mainly because it isn't exactly clear how some physical base of a mental property would in turn depend on it's supervening mental property.

While this may seem better territory than token physicalism to create a theory of how the mind and the body are related, supervenience itself is not enough. It, like token physicalism, is a shared commitment of various mutually exclusive mind-body theories. For example, mind-body supervenience is a component of our emergentist theory, but it is also component of reductive physicalism. Thus, mind-body supervenience cannot itself be a fully articulated position in the mind-body problem, precisely because it is a shared commitment of conflicting theories of how the mind relates to the body.

Further, mind-body supervenience is not a metaphysical thesis; it merely states a pattern of property *covariation*. The additional question still remains as to why the supervenience relation between the mental and physical holds. Is it because the mental is simply reduced to the physical? Or is the supervenience relation something that must be accepted as a brute fact of nature? Supervenience itself cannot answer the question as to

what *grounds* or *accounts* for the relation. In this sense, mind-body supervenience is another way to state the mind-body problem if one holds a determination relation between the mental and physical. It is not itself a solution to it (Ibid., p. 14).

As formulated above, supervenience is a common characteristic of all basically physicalistic theories of how the mind and body are related. Consequently, supervenience has been defined as *minimal physicalism* (Kim 2000, Lewis 1983). "Minimal" here is used in the sense that any theory that hopes to be physicalistic must accept supervenience as one of its doctrines; it is not minimal in the sense that it alone can push some prospective mind-body theory inside the borders of physicalism for the reasons outlined above. Supervenience represents the idea that mentality *at bottom* must be physically based, and it cannot float freely and unconnected to the physical nature of the objects in which it is instantiated (Kim, 2000, p. 15). The physicalist is obviously allowed to hold a more reductive view of mentality in the sense that mentality is *entirely* physically based, but the scope of physicalism lies within the parameters of supervenience and reduction.

4.3 Physical Realization

We know now that an appeal to mind-body supervenience alone is not enough to constitute a solution to the mind-body problem, so moving forward our task is to find an additional requirement that closes the explanatory gap. Jaegwon Kim, among others, have suggested that the addition of the idea that mental properties are *physical realized* to mind-body supervenience is able to do precisely this. Thus, I will examine physical realization here.

Physical realization- mental properties, if they are realized, must be physically realized- that is, no mental properties can have nonphysical realizations (Ibid., p. 19).

This is equivalent to the conjunction of physicalism applied to the functionalist conception of mental properties:

Functionalism- mental properties and kinds are functional properties, properties specified in terms of their roles as causal intermediaries between sensory inputs and behavioral outputs (Ibid., p.19).

When functionalism is understood in a physicalist form, the only potential occupants or *realizers* of the causal roles alluded to above are physical properties. For example, the functionalist asserts for something *X* to be in pain is for *X* to be in some internal state that is caused by tissue damage and further causes characteristic pain behavior such as wincing or crying. This then means that the mental state of being in pain is a *second-order property*. Second-order properties are defined as follows:

> F is a second-order property over set B of base (or first-order) properties if and only if F is the property of having some property P in B such that D(P), where D specifies a condition on members of B (Kim, 1997, p. 280).

If we consider the example of pain again, D then specifies some certain causal role characteristic of being in pain, and that is tissue damage that results in wincing or crying. To be in pain, then, is to have some physical property fill the causal role of being in pain. For humans, this physical property is the firing of C-fibers, since the firing of C-fibers fills the causal role specified by D. The base (or physical) properties that satisfy condition D are

called the *realizers* of the second order property F. Thus, we can say that the property of C-fibers firing physically realizes the second-order mental property of being in pain.

When the principle of physical realization is added to the mind-body supervenience thesis, we are then in a position to explain why the supervenience relation holds. According to physical realizationism, the mental supervenes on the physical because mental properties are second-order functional properties with physical realizers and no nonphysical realizers (Ibid., p. 283). When we wonder why whenever C-fiber are fired in humans the mental property of pain is instantiated, we have an answer. For a human to be in pain, this simply amounts to C-fibers firing; or, as some like to say, for a human to be in pain, there is nothing "over and above" the firing of C-fibers.

On this account, we now have the tools to respond to the question of how the mind is related to the body. This is because we now have an explanation why the supervenience relation between the mental and the physical holds- it is due to physical realization.[8] The firing of C-fibers is always accompanied by the sensation of pain, because the sensation of pain is a second-order property that is physically realized by the firing of C-fibers. Further, if a mental property is physically realized, and if we have full knowledge of the realizing property as well as the accompanying laws of nature, it is then possible to deduce all the characteristics of the mental property in question. The mental property of pain has been reconstrued functionally, and, further, we have an explanation how C-fibers fill this causal role. If we consider the type of functional analysis explained in section 3.4, this type of analysis amounts to a reductive explanation of mental properties.

---

[8] To put this more carefully, physical realization provides a possible way to close the explanatory gap. Whether or not it actually holds is obviously a point of debate.

The emergentist though, asserts that mental properties[9] are irreducible. By this, we can now assume the emergentist mean that mental properties are not *functionally analyzable*. This means that emergent properties *do not count as physically realized properties*. The emergentist asserts that when some system instantiates P, mental property M emerges (in some nonphysical sense) and that it is not a physically realized property. For the emergentist to explain the supervenience relation, they must rely on something like psychophysical laws rather than physical realization. From here, it should now be clear that emergentism does not qualify as a form of physicalism if these stronger principles of physical realization are required to fully capture what is at the heart of a physicalist theory.

4.4 Emergentism as a Form of Dualism

While emergentism is consistent with token physicalism and supervenience relations, it does not seem that either is enough to capture physicalism. On the contrary, stronger forms of physical realization do indeed seem to be needed to guarantee physicalism, but these forms of physical realization deny that emergent properties can be part of their scheme; so, in this sense, emergentism is at odds with physicalism. I believe that this argument shows that emergentism is better viewed as something closer to a form of dualism.

This is then the short answer to our first question of whether or not emergentism is a form of physicalism or dualism. It may seem because emergentism now has a dualist label, the emergentists were not successful in their original goal of inhabiting a genuinely new middle space between the extremes of vitalism and mechanism, but I believe this question is

[9] At the very least, phenomenal properties of consciousness.

far more nuanced than we might first imagine. To examine why this may be the case, it will be necessary to further examine emergentism as a dualist theory.

First, the emergentist is committed to physicalism monism, and this is captured by requirement *i* of our definition of an emergent property constructed in the last section. For a reminder, requirement *i* states that an emergent property is composed of physical elements completely from the lower level. This commitment alone means that emergentism is in conflict with stronger forms of dualism of the Cartesian sort. Property dualism is a weaker form of dualism that is the conjunction of the acceptance of physical monism and the rejection of the claim that all properties possessed by physical systems are physical properties (Kim, 2006a, p. 290). Further, property dualists are also mental realists in that they regard mental properties as genuine properties that can make a causal difference on the physical world.

Emergentism is also committed to these three claims, which are highlighted as requirements *i*, *iii*, and *v* from our definition of an emergent property. In this sense, emergence can be viewed *as a form of property dualism*. While the emergentist and the property dualist agree on these three claims, there is a very important consideration relative to which the two diverge, and that is the supervenience thesis. The emergentist goes far beyond many forms of property dualism by accepting a very intimate connection between mental properties and physical properties. The property dualists, on the other hand, need not to accept the claim that the mental supervenes on the physical in any way, because they need not posit any dependence relation of the mental on the physical.

By not requiring any dependence relation between the mental and the physical, it appears that the property dualist's conception of how the mental and the physical are related

seems prima facie scientifically unacceptable. As we have seen in the section prior, supervenience has been defined by Kim as *minimal physicalism*, in the sense that it is a shared commitment across all physicalist theories. Given that emergentism accepts mind-body supervenience, I believe that this commitment makes emergentism *the most interesting form of dualism*. By most interesting, I mean that it holds an essential component of any physicalist theory, but it also synthesizes this with an irreducibility claim, which is an essential component of any dualist theory.[10] In this sense, emergentism is a true hybrid theory. By accepting the supervenience thesis, emergentism aims to make mental properties subject to empirical scientific investigation. For this to be the case, the emergentist can appeal to nomological psychophysical laws that explain the connection between mental and physical properties.[11] The property dualist position as generally stated does not seem to share this requirement. This then makes emergentism the weakest form of dualism in the sense that it places restrictions on the mental, but logically stronger than other forms of property dualism because it has an additional requirement of mind-body supervenience.

If we interpret the goal of the emergentist to create a middle ground between mechanism and vitalism as simply equating to the goal of creating a truly nonreductive physicalist theory,[12] then it seems that much of their effort has been in vain. I do not believe this is the best interpretation, though. For one, it may be the case that *no nonreductive physicalist position at all is stable*, and each is in danger of collapsing into either reductive

---

[10] Interactionism also shares these characteristics, but for our purposes, the two are essentially synonymous. We can shortly say it's more interesting than epiphenomenalism and parallelism, because they have no room for the causal efficacy of the mental.
[11] Of course, the idea of the existence of such psychophysical laws themselves is a matter of controversy, and this issue will be taken up later.
[12] By this, I mean a theory that can rightly be called physicalist while still holding onto the irreducibility of the mental.

physicalism or more serious forms of dualism. If we assume the functional model of reduction, functionalism looks something very like reductive physicalism, whereas the functional model renders emergentism as a form of dualism.[13] Thus, it may be that this goal can never be accomplished.

As stated before, there is both a positive and negative element to the emergentist goal as I have formulated it above. It does seem that the emergentist was successful in creating a theory that does not result in a complete reduction of mental phenomena without positing any supernatural elements into his or her theory, but property dualism functions in the same way, so these elements alone cannot be enough for success. When we consider the positive elements of the emergentist goal, it is not exactly clear as to whether or not emergentism adequately meets both standards. It is fairly uncontroversial to say the emergentist has indeed saved a special place for mental properties within the framework of their theory, but as for whether or not they do so in a scientifically acceptable way is a much more complicated question.

The supervenience thesis was intended to be a way for the emergentist to accept the intuitive physicalist claim that the mental must in some way be dependent on its physical structure. We also know that the type of ontological emergence that is being presented within this paper involves a kind of mental realism that asserts that mental properties have a genuine causal influence within the world. For the latter proposition to be the case, it seems that the physical must also depend on the mental in some way. While this type of converse relation of determination between the mental and the physical is allowed within the doctrine

---

[13] For a more detailed argument on this point, see Kim "The Myth of Nonreductive Physicalism" In The Philosophy of Mind ed. Brian Beakley and Peter Ludlow (Cambridge: The MIT Press, 2006b) 427-442.

of supervenience, there have been objections to the idea that this kind of relationship can actually exist in our world. For example, Jaegwon Kim's (1989, 1990, 1993, and elsewhere) exclusion argument suggests that, for nonphysical mental properties to causally influence physical properties, one has to deny the causal closure of the physical. The causal closure of the physical states: we do not need to look outside the physical domain when searching for some physical event's cause. If this principle were not the case, we could never have a complete theory of physics without invoking non-physical causal powers. It seems that the emergentist is committed to something very close to this, but this clearly flies in the face of our scientific understanding of the world.

The explanation of how this could be so is just one of the problems the emergentist must face when considering mental causation. Mental causation has been a metaphysical problem the dualist has constantly been reminded of, so it should not be surprising it also applies to the emergentist given my argument above suggests emergentism to be a form of dualism. Thus, for the emergentist to be able to declare success, they must first resolve the problem of mental causation in a scientifically respectable way.

5. The Exclusion Argument

Before we can understand how the emergentist may go about solving the exclusion problem, we have to become a little more acquainted with the exclusion problem itself. Anyone acquainted with the literature of philosophy of mind is also most likely very familiar with the structure of the exclusion problem. This is probably for two reasons. First, it is an argument by dilemma, and its consequences are unacceptable for a proponent of just about every non-reductive theory[14] of the mental who is also a mental realist. Second, it is a very good argument that allows for no clear or obvious way out for the non-reductive mental realist. Thus, its popularity in the literature is due to both its importance and its difficulty for those who aim to solve it. The first formulation of the exclusion argument can be credited to Stephen Pepper (1926), but its most famous advocator is most definitely Kim. He is credited for much of its more sophisticated formulation, and also applying it to the current debates within philosophy of mind. Kim has brought up the argument in many of his works, but here I will focus on an account from *Philosophy of Mind* (2006).

The exclusion argument arises out of a commitment shared among most physicalists- that the physical domain is causally closed. If we trace the history of any physical event, we need not go outside the boundaries of the physical domain to explain the event. Expanding on this point, every causal chain concerning physical events needs not to cross from the physical over to the nonphysical. The principle can be stated as follow:

Causal Closure of the Physical Domain- If a physical event has a cause (occurring) at time t, it has a sufficient physical cause at t (Kim, 2006, p. 195).

[14] The only exception to this may be dualist parallelism.

First, I should point a few things about this view. It does not explicitly state that there can be no nonphysical cause for some physical event. Rather, all it says is that when we search for the physical event's cause, we need not to look outside the domain of the physical. Second, it does not imply a type of physical determinism, because it does not state that every physical event has a cause or causal explanation. Third, formulated as such, it is not directly inconsistent with dualism. The principle does not conflict with the existence of immaterial souls. It is only inconsistent with the claim that the nonphysical souls inject causal influence into the physical domain.

What would it mean if the closure principle didn't hold? If it didn't hold, there would be physical events for whose explanations we would have to look to nonphysical causal agents. Such causal agents could be something like souls, angels, spirits, divine forces, or anything else outside the fabric of space-time. In this sense, the closure principle does conflict with Cartesian interaction dualism. Obviously, the postulation of nonphysical causal agents is unacceptable to any physicalist or, for that matter, to anyone who wishes to hold a naturalistic standpoint. Further, if physical closure did not hold, theoretical physics would be in principle *incompletable*. Surely, it seems as though physicists operate under the assumption that physics is a complete science.

It may be a little clearer now as to why the closure principle is not kind to mental causation- particularly mental-to-physical causation. Suppose that mental event *M*, causes a physical event, *P*. Given that the emergentist is committed to mind-body supervenience, there must also be a physical cause of *P*. Call it event *P**. *P** is then the supervenience base for *M*, thus it occurs synchronically with *M*. Further, according to the closure principle, *P** is the sufficient cause *P*. Now here is where the dilemma arises.

This first option is to say that $M=P^*$, thus identifying the mental cause with the physical cause as single event. This then turns the supposed case of mental-to-physical causation into an instance of physical-to-physical causation. There is no mystery of causation here, but it's clearly an instance of reduction, thus it goes against the emergentist principle of the irreducibility of the mental. Further, we also have reason to reject the principle because of our knowledge of multiple-realizability of the mental by the physical.

Thus, the emergentist says there can be no reduction of *M* to *P**. The next option is to claim event *P* had two distinct causes- namely *M* and *P**. This then leads to *causal overdetermination*, or that one event has two, distinct and sufficient causes. If we accept this horn of the dilemma, then we are committed to the idea that every instance of mental-to-physical causation is an instance of causal overdetermination. A corollary commitment can be stated as that if the mental cause had not occurred, the respective physical cause would have still brought about the physical effect. There are two problems here.

First, this appears to be an absurd conclusion. To understand the absurdity, it is useful to first imagine a case of isolated causal overdetermination. Imagine the textbook example of the firing squad, in which two executioners fire their respective weapons, propelling two bullets toward the heart of the executionee. Both bullets pierce the heart at precisely the same time. Thus, both shots were sufficient causes for bringing about the effect of death of the executionee; and if for some reason, one of the shots had not occurred, the other shot would have still brought about the effect of death. Isolated incidents of causal overdetermination themselves are not metaphysically unacceptable. We can accept this kind of situation, because we consider the case above to be rare and the exception to the rule, rather than the precedent. If mental-to-physical causation is always a

case of causal overdetermination, then something like the firing squad example is happening everywhere and every time there are mental subjects. While we can accept isolated incidents of causal overdetermination, widespread and systematic causal overdetermination seems bizarre at the least and unacceptable at most.

The second problem is that causal overdetermination seems to weaken the status of the mental cause. For *M* to be a genuine cause of some physical event, it relies on a synchronous physical event *P* that also serves a sufficient cause of the physical event if we assume the mental supervenes on the physical as the emergentist does. It seems that for *M* to be a genuine cause, it should be able to bring about some physical event without the help of some additional sufficient physical cause. If we take all of this together, every mental event has an accompanying physical cause that is sufficient to bring about the effect even if the mental event had not occurred.

The picture presented by this causal competition then leads to an additional principle that is not particularly kind to mental causation either. It can be referred to as the *Exclusion Principle*:

Exclusion Principle- No event can have two or more distinct sufficient causes, all occurring at the same time, unless it is a genuine case of overdetermination (Ibid. p. 196).

The example of the firing squad would qualify as a genuine case of overdetermination. To state the principle more generally, genuine causal overdetermination occurs when two independent causal changes converge at a single effect (Ibid. p. 196). With all this in place, let's return to our case of mental-to-physical causation. Let's again begin with the general assumption that there is a mental event that causes some physical event such that:

(1) *M* is a cause of *P*

Now, according to physical causal closure, it follows that there is also a physical event *P**:

(2) *P** is a cause of *P*

If we apply the irreducibility principle included in our emergentist theory, we can make the additional claim that:

(3) $M \neq P^*$

Now, suppose we also assume that:

(4) This is not a case of overdetermination

When we consider both the closure principle and the exclusion principle in conjunction with these four claims, we run into some serious trouble. Claims (1)-(3) together amount to the proposition that the effect *P* has two distinct causes- namely *M* and *P**. (4) states this is not a case of causal overdetermination, thus bringing the exclusion principle onto the scene. One of the causes, either *M* or *P**, must be eradicated as the cause of *P*. Now comes the closure principle. The closure principle states there must be a physical cause of *P*, which is *P**, so consequently *M* is eliminated from the causal scene, and *P** does all the causal work.

There is a way to save mental causation in the face of the exclusion argument. If we concede that $M = P^*$, then there is only a single event made up of one cause, thus there is no competition. This would amount to embracing mind-body identity, but it does not seem that this option is open to the emergentist, given their anti-reductive commitments. If the emergentist maintains that $M \neq P^*$, then the mental is epiphenomenal, which may be enough

for some, but not for the kind of emergentist that is being presented in this paper. This kind of emergentist is a mental realist, and asserts that the mental must have some kind of causal efficacy for it to exist in any meaningful way. Thus, the emergentist is caught in a serious dilemma, both horns of which deeply threaten the plausibility of her view.

6. A Response to the Exclusion Argument

There are a couple of different ways to go about attempting to solve the exclusion problem.[15] Option (1) concerns the claim that mental and the physical are distinct. One could concede that the mental and the physical are not actually distinct, which would amount to some kind of reductive physicalist thesis. Option (2) deals with the exclusion principle. Once again, the exclusion principle states, "no single event can have more than one sufficient cause occurring at any given time- unless it is a case of overdetermination" (Kim, 2006, p. 39). To be successful in this kind of approach, a) one can argue that mental/physical overdetermination does indeed occur, and that it is pervasive, or b) one could argue that some physical effects can have mental causes, but they are not overdetermined. Option (3) focuses on the causal closure problem. To be successful in this endeavor, one could argue that the physical world isn't causally closed, thus leaving room for mental properties to exert their causal influence 'downward.'

Many attempts have been made at all of these possible solutions by a wide range of philosophers, and some have been more successful than others. Even though there are many options of how to go about solving the causal exclusion problem, I am not after just any possible solution. Rather, what I am after here is the best possible response *for the emergentist given the theoretical framework that has been presented thus far*. Recall requirement v) of my definition of an emergent property: requirement v) states that emergent property P has causal powers of its own and these powers are irreducible to the causal powers of its basal constituents. Also recall that the emergentist wants to claim emergent properties are *genuinely*

[15] The list that follows is not comprehensive, but it will be complete enough for our concerns here.

novel, and novel is used in the metaphysical rather than epistemological sense. Robert Van Gulick states that a metaphysical definition of emergence can focus on either properties or causal powers (2001, p. 17). I believe that for the emergentist to answer to the exclusion problem, it is best to focus on causal powers for two reasons. The first is that it seems as though the emergentist is committed to the claim that emergent properties have causal efficacy given requirement v) of our definition. If the emergentist cannot clearly explain how emergent properties exert their causal influence to lower-level base components, then we have good reason to question the emergentist thesis. Thus, emergentism stands or falls with either the success or failure of the explanation of how an emergent property can exert causal influence. Nancey Murphy has even gone as far as saying "emergence needs to be defined in terms of the denial of causal reductionism" (Murphy, 2006, p. 227). The second reason why it is best to focus on the causal powers of emergent properties is because some property having some kind of causal power not possessed by its corresponding base property seems to amount to that property being genuinely novel in the sense described in section 4.

Given that the irreducibility of the mental to the physical is a characteristic commitment of any emergentist thesis, option (1) of solving the exclusion problem above is clearly not available for the emergentist. Option (2) is an option open to the emergentist, but many attempts in this vein focus heavily on causal explanation, rather than directly arguing that emergent properties possess their own kind of causal powers. For this reason, option (2) does not seem to yield the theoretical implications the emergentist is committed to. This then leaves option (3), and some (Kim, 2008) believe it to be a natural step for the emergentist to take. This is not a very popular position to take, though, because it directly undermines the possibility of a complete theoretical physical science. Physics would have to

incorporate various psychological properties into its calculations for it to able to sufficiently explain the world. This formulation clearly does not jibe with the naturalistic standpoint that is aimed at here. As stated above, though, the emergentist had better offer some sort of explanation of an account of the efficacy of emergent mental properties. This will generally require some form of downward causation, but, given our naturalistic commitments, our account of mental efficacy cannot violate the closure principle. This may seem an impossible task, because, generally, the affirmation of downward causation is thought to equate to the denial of the causal closure of the physical, but some philosophers have suggested that this isn't always the case. For this claim to hold, one must present a special kind of downward causation of mental properties that does not disrupt causation at lower levels. Further, given our commitment to requirement v) the causal powers of the mental must be irreducible causal powers. In what follows, I will examine whether all of these claims are consistent with each other.

6.1 Top-Down Causation

One such example of the kind of argument highlighted above comes from Nancey Murphy in *Emergence and Mental Causation* (2006). She suggests downward causation can be defined by *selection* among lower-level processes on the basis of the higher-level supervenient properties. Further these higher-level properties have an irreducible causal role to play, and it is only in virtue of the higher-level mental properties that the lower-level neural processes become subject to selective pressures of the environment (Ibid. p. 227). Her reasoning then invokes some kind of *externalism*. Externalism, stated roughly, is the idea that the mind is not simply the result of what is going on in the brain (or the nervous system, more extensively), but it also requires an appeal to either what is going on or exists *outside* of the brain. This

then involves *intentional* mental properties. Some examples of intentional mental properties are propositional attitudes, mental images, and perceptual experiences. These kinds of properties are *about* something, in that they are mental states that have representational content. Externalism then assumes that the content of representational states depend on *relations* between the person and the environment, not merely on the intrinsic features of the representational states. The relations of particular interest are the historical, social, and causal relations the person bears to their outside environment.

A classical example for the argument of externalism about mental content comes from a thought experiment put forth by Hilary Putnam (1975).[16] Consider Oscar, who lives here on earth, and Twin Oscar, who lives on Twin Earth. On Twin Earth, there is no H_2O, but rather, an indistinguishable liquid composed of molecules XYZ. Call this liquid "twater." Further, suppose that Oscar and Twin Oscar are indistinguishable in their intrinsic make-up. Now imagine Oscar expressed the belief that "water quenches thirst." This belief can only be true if H_2O quenches thirst. Because Twin Oscar has never encountered or heard of H_20 (he only knows about twater), Twin Oscar does not believe that H_20 quenches thirst. When he expresses the belief that "water quenches thirst," he was expressing a different belief: the belief that twater quenches thirst. This belief then has different truth conditions from the belief the Oscar expressed. Thus, despite being intrinsically indistinguishable, Oscar and Twin Oscar have different beliefs. The externalist argues what follows then is that meaningful states owe their meaning to more than just intrinsic make-up alone. Thus these beliefs are *broad content*, because they depend on both features of the environment and features of the individual.

---

[16] His original argument was not aimed at mental content. Rather, he applied it to linguistic content, but his thought experiment has been widely applied to mental content by others.

Her account is also highly influenced by Fred Dretske's (1988) distinction between *structuring* and *triggering* causes. A *triggering* cause can be thought of as the event that initiates (or "triggers") some behavioral process that ends in some sort of movement. The triggering cause causes this process purely in virtue of its intrinsic properties. The *structuring* cause, in contrast, amount to the events that *shaped* or *structured* the behavioral process. The structuring cause is what caused *a* to cause *b*, rather than causing *c*. In this sense, the structuring cause is responsible for it being *this process* of *a* causing *b*, rather than some other process (Dretske, 1993, p. 123).

The claim is then that, although the properties picked out by psychology and the other special sciences are indeed made up completely of aggregations of lower-level physical components, the causal powers of an object are not entirely determined by the physical properties of its constituents, but also by the *organization* of those constituents within the composite. Further, the patterns of background conditions picked out by psychology have a downward causal efficacy not in the sense that they can cause any part of the microphysical state of affairs, but they have efficacy in that they can affect which causal powers of the micro-constituents are *activated* or *likely to be activated*. Thus, we can agree with Kim that synchronic reflexive causation is likely false. *Synchronic causation* requires that a supervenient property must alter a lower-level property the instance it is realized, and *reflexive causation* states that a higher-order property causes a change in its own base property. It cannot be the case that an emergent property at time t causes any part of the microphysical state of affairs that constitutes the instantiation at t of the corresponding physical micro-structural property (Murphy, 2006, p. 35).

Murphy, like any good emergentist, claims that the mental supervenes on the physical, but she explains the supervenience relation in a non-standard way. Her definition of supervenience states: "property S supervenes on (base) property B if and only if entity e possesses S in virtues of e's possessing B under circumstances *c*" (Ibid., p. 230). Her definition of supervenience is a little more metaphysically open than what has been presented thus far in the sense that, her definition of supervenience allows her to say that mental properties supervene on brain properties, but at the same time, she can also say that some brain properties are co-determined by environmental factors. The kind of supervenience she is invoking her is known as *global supervenience*. The kind of supervenience that has been discussed thus far assumes *local supervenience*. Global supervenience can be stated as such: for any property S of individuals, any two possible worlds indistinguishable in physical respects at time t will agree in the instantiation of S at t. Strong local supervenience can be stated as follows: for any macro property S on an individual, there is some physical property (or a set of some combination of physical properties) of the parts of the individual, such that, it is nomologically necessary that any individual instantiating any member of that set at time t instantiates S at t (Berent, 1995, p. 173). So, to put this in a more technical sense, the locally supervenient mental properties of the individual in question are associated with certain globally supervenient properties, and the globally supervenient properties will have their base extended to a region outside the individual in which properties causally interact with the locally supervenient properties of the individual in question (Ibid., p. 175). While this diverges from the standard account in some important ways, the value of this definition is that it may allow the emergentist a way out of the exclusion problem, because, as stated thus far, the exclusion problem deals only with the locally supervenient properties.

I will not attempt to refute her use of this kind of supervenience here, but some concerns will surface later.

With this type of supervenience relation, the standard picture of the exclusion problem is too isolated of a model for there to be any kind of mental causation, because it represents too little of the causal history, thus it must be expanded. Murphy gives an example of a man, call him John Canine, who happens to be in prison. Canine has learnt through experience that when a bell sounds, the cells are unlocked to allow the men out for their meals. Under these conditions, let them be C_1, when Canine hears the bell, he pushes on his door, and proceeds to get lunch. The situation can be depicted as such:

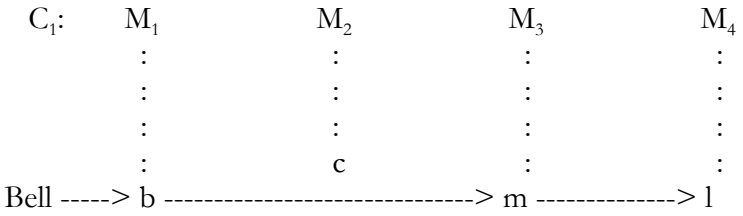

Fig. 10.7 (Murphy, 2006, p. 237)

Let M_1 represent Canine's hearing the bell, M_2 represent his belief that the door will open when the bell rings, M_3 represent his conscious decision to open the door, and M_4 represent him enjoying lunch. Further, let *b* represent the neural correlate of Canine hearing the bell, *c* the cell assembly that is the neural realization of his belief that the door will open when the bell rings, *m* the set of events in the motor cortex that initiates the door-opening, and *l* his eating lunch. Also, let the vertical dotted lines represent supervenience relations and the horizontal arrows represent causal relations.

Some more needs to be said about *c* and M_2 before we can attempt to explain a causal interaction from this diagram. We are not so much interested in the supervenience relations connecting M_1 to *b* or M_3 to *m*. Murphy is interested in the newly formed

connection between *b* and *m* that is represented by *c*. For simplicity's sake, just imagine that there happens to be a single neuron connecting *b* and *m*, and the causal line connecting the two depicts it. Prior to the conditions stated in C_1, the connection *c* was not in place (or at least not strongly connected), but now that *c* is in place, it is a *bearer of information* about Canine's past environment- namely of the relationship of bells and doors opening. She then claims, "*Canine's brain has acquired a new supervenient property, M_2, the property of being a representation of the relationship between bells and doors opening*" (Emphasis in original, Ibid. p. 233).

There are some immediate objections that need to be attended to before we can proceed to the causal story. Murphy claims this is clearly an instance of downward causation that explains the existence of the newly formed connection referred to as *c*. The neuron *b* would've been multiply connected to various regions of the brain, but there pairing of *b* with *m* resulted in the *selection* of the newly strengthened connection. This is because the newly formed neural connection is embedded in the broader causal system of Canine's history of bells and doors. If it were considered apart from Canine's causal history, it would be just another neuron, and it would not escape the grips of the exclusion argument. Because Murphy's definition of supervenience allows some mental properties to be co-determined by environmental conditions, she is then able to insist that the existence of c *is due also* to broader causal system's circumstances by which it was formed.

Despite this move, it still appears that M_2 is merely epiphenomenal. If the belief were not there, things would proceed in just the same way, and the causal explanation needs only the physical facts of the situation. Murphy is not trying to claim that the causal arrows should be drawn from *b* up to M_2 and then from M_2 back down to *m*. Remember, she does not want to violate the causal closure principle. Of course, in epistemological terms, the

existence of M_2 is explanatorily relevant for why the connection is there, but not causally relevant for explaining how hearing the bell leads to Canine's getting lunch.

Murphy recognizes this obvious difficulty, and she suggests for there to be a causal role for M_2, we have to consider the future. For Canine to change his behavior, he can alter the neuronal connection by means of his belief's representational relationship to the environment. Thus, the supervenient representational property is the 'handle' that allows for selective causal impact on the neuronal connection by means of the environment (Ibid. p. 235). Now, imagine that the situation changes from C_1 to C_2, where the bell and the lock are no longer synchronized. It can be depicted as follows:

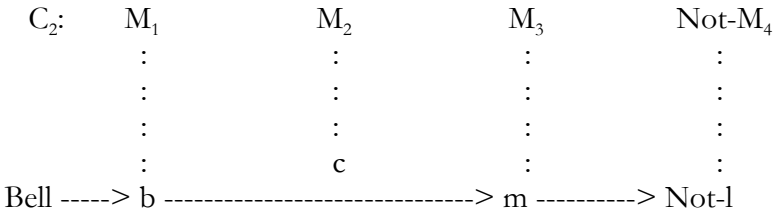

Figure 10.8 (Ibid. p. 238)

Once again, M_2 does no causal work in bringing about *m*; *b* and *c* are sufficient for *m*. Murphy also recognizes this, but claims that, however, the relationship between the environmental circumstances (C_1) and the representational belief (M_2) is paramount for the final outcome of getting lunch or not. The representational belief is qualitatively the same in either situation, but in C_2, M_2's relation to the world has changed *and it no longer has the representational character it had before*. What was a true belief before has now become a false

belief,[17] and as such "*its neural realization c has a different effect in the world than before"* (Emphasis in original Ibid. p. 237).

Now, even if we eliminate all the supervenient properties, *b* and *c* together will still produce the same immediate causal effect. To pull this kind of situation apart from the causal picture Murphy is trying to paint, she asks us to imagine that Canine has been hypnotized, such that he is not able to consciously hear the bell, and thus not be able to form the belief that when the bell rings the door unlocks. Canine will not be able to evaluate his beliefs in this scenario because he has none relating bells and doors, and thus he will not be able to change his behavior appropriately *because he has no access to the neural processes that cause his behavior*. In the normal condition when Canine is not hypnotized, he can quickly learn that his belief is now false; or, he is able to evaluate M_2, and thereby evaluate his neural connection *c*. Thus, Murphy concludes that "the conscious mental property is causally relevant to the subject himself when there is need for a change" (Ibid. p. 238).

6.2 Objections and Remarks

While it may be thought that the emergentist makes up some ground by this kind of argument against the exclusion problem, I have a couple of concerns about the story being told above. First, it is not clear that this is the kind of causal relevance we are normally after. On this account, part of the causal explanation is the selection of certain neural pathways by means of downward causation. Let us accept that this can actually occur for argument's sake. This part of the causal explanation then requires facts from past learning histories. The story Murphy told above about Canine doesn't vividly raise this concern because the

[17] This once again goes against the standard account of supervenience, but it's advantageous in the sense that she can now agree with the externalists of mental content.

temporal period between C_1 and C_2 appears to be relatively short, but imagine this: Canine has been in prison for 30 years, and every day from the beginning of his sentence, when the bell rings it has been accompanied by an unlocked door, represented by C_1. Today, when we ask why he opened the door when the bell rang, our explanation must refer to the original selection of neural pathway *c* that occurred 30 years ago according to Murphy's account. This is because that was when the relevant selection of that neural pathway occurred. This strikes me as implausible at most, and unintuitive at least. When we explain actions of behavior by mental states, we generally regard mental properties as causally relevant to what is going on at the very moment of the behavior.

Murphy may very well respond by saying that this type of causal relevance of mental properties is all we can hope for, and that it is still enough to save some kind of minimal mental causation. Further, the only reason why we don't find it satisfying is because it is not our intuitive account of the mental's causal relevance. I'm fairly sympathetic to this response, mainly because I find it very plausible that our intuitions about mental causation are mistaken, but not necessarily in the way Murphy might think they are. Nevertheless, I will not rest my critique of Murphy's account of mental causation on this point, because I believe her account faces a much greater difficulty than the one pointed out here.

An additional worry of mine is whether or not selection and changing of behavior via mental representations play any important causal role in the way Murphy seems to think they do. First, let us consider selection. Murphy claims selection of relevant neural pathways occurs via downward causation of the system interacting with its environment, and that the presence of the new supervenient representation strengthens that connection. At this point, we have to ask ourselves if this really solves the exclusion problem. She claims *c*

is a bearer of information about Canine's environment- namely the connection between bells and doors. Because *c* is a bearer of information, Canine acquires the new supervenient property mental property M_2- the property of being the representation of the relationship between bells and doors. She claims that, because her definition of supervenience allows for an appeal to globally supervenient properties, this is the only way *c* can bear information. She claims if we consider the standard count of supervenience, the information-bearing capacity of the new mental property falls away. She justifies this claim by saying that the information is dependent on the existence of *c*, but also on the circumstances under which it was formed (Ibid., 234).

While this move may seem to make M_2 an indispensable component of the causal explanation, we have to ask ourselves the Kimian-style question of what's doing all the work here? Once again, Murphy claims that the existence of *c* is due to an instance of "clear" downward causation (Ibid., p. 234), but this is a very weak notion of downward causation that give *no causal role to any mental properties*. The underlying neurophysiology is completely sufficient for the formation of the new neural connection *c*. While Murphy argues that the circumstances under which it was formed matter, it seems contrarily that Canine's brain would only be sensitive to the local features, and they would *screen off* any of the possible external factors. Thus, the external factors have no causal relevance to the wiring of Canine's brain. While Murphy calls this an instance of downward causation, it does not seem to be an instance of downward causation that bears any causal relevance. Any kind of causal relevance of the higher-level properties is excluded from the causal picture, so the higher-level properties are still epiphenomenal.

In defense of Murphy, she does not rest her causal story of mental properties on this instance of downward causation; she claims there is additional causal relevance for the supervenient mental property for M_2 to play when Canine is in a situation in which he must change his behavior due to changing circumstances. This is because, by virtue of M_2, Canine is able to evaluate his neural connection *c* and change it accordingly. While M_2 may be explanatorily relevant in this sense, *it still does not seem that the supervenient mental property is doing any causal work*. It seems rather that the physical properties that underlie M_2 are the properties that do all the relevant causal work. For essentially the same reasons cited above, all the external goings-on are screened off by the local factors of the brain, and we are left with a possible mental cause and a possible physical cause. If we apply exclusion reasoning, one must be discounted from the causal picture. If we then apply the causal closure principle, it is clear that the mental cause gets dispensed and the physical cause stays. Thus, once again, the mental cause is epiphenomenal.

7. Concluding Remarks

The goal of this paper was two-fold. As a reminder, the first was to make sense of the concept of emergence, and apply it to philosophy of mind. In the sections that dealt with this concern, I constructed what I believe is a definition of both necessary and sufficient conditions for what is required for some property to qualify as an emergent property. The most important part of this definition concerns the irreducibility claim put forth by the emergentist. Within the irreducibility claim, there are four separate claims that make emergentism a distinctive mind-body theory. The first is the supervenience claim. This claim is important to the emergentist because it posits a tight connection between the mental and the physical. Because of the tight connection between the two, we have a way to describe aspects of the mental in terms of the physical, thus allowing for empirical investigation. The supervenience claim then satisfies the emergentist's hope for a naturalistic standpoint. Another related claim the emergentist is committed to is the existence of irreducible psycho-physical laws. This is an empirical issue for the emergentist.

The second claim of importance within the irreducibility claim is that emergent properties are not reductively explainable in terms of its basal constituents. We know that from preceding discussion the relevant type of reduction is *functional reduction*. If this is truly the appropriate mode of reduction (which is a claim that I believe), then it really does seem that at least phenomenal consciousness is truly irreducible. We do not define phenomenal consciousness in terms of its functional role, but in this sense, the lack of a functional role for phenomenal consciousness is a double-edged sword. Given that the kind of emergentist presented in this paper is committed to the fourth claim that an emergent property has irreducible causal powers of its own, a lack of a functional role for phenomenal

consciousness seems to put the causal efficacy claim in jeopardy (at least for the efficacy of phenomenal consciousness).

With all these claims in place, I then presented an argument that concluded emergence is a dualist theory, rather than a physicalist theory. This is a fairly significant point, because emergence is generally considered one of the seminal non-reductive physicalist theories. It is also a significant finding for emergentists, because if emergentism is a dualist theory, then the emergentist must deal with the same kind of problems the dualist must confront. The most imminent issue for the emergentist is then the problem of mental causation.

The second part of the goal was to assess the tenability of emergentism as a mind-body theory. It should be clear that the tenability of emergentism hinges quite heavily on the debate of mental causation as formulated by Kim with his exclusion argument. From my arguments in section 6, I conclude that the emergentist still has a lot of work to do, and many of the prospects are quite bleak in the face of exclusion. Essentially, the claim Murphy was resting on in her debate for emergent mental causation, was the causal efficacy of broad mental content. While this may seem a good way to ground some kind of mental causation because seems to add an essential role for the mental in the sense that it connects intrinsic features of the brain to its outside environment, it does not seem to yield any causal efficacy for the mental. It still falls victim to the exclusion argument. Of course, this is only one notion of broad mental contents, but it seems to me that any attempt to explain the causal efficacy of broad mental content will be screened off by the local, intrinsic features of the brain.

It should also be noted that this is only one possible way to attempt to solve the exclusion problem, so this failure of this attempt does not mean the death of emergentism. The reason why I chose Murphy's argument was because it directly argues for the causal efficacy of mental properties within the emergentist framework. Many other attempts to solve the exclusion problem focus more closely on accounts of explanatory relevance (Horgan, 2001, Yablo, 1992, 1997). These types of arguments do not directly argue for the causal efficacy of emergent mental properties, but if they work against the exclusion problem, so much the better for the emergentist. Their task is then to argue that these more explanatory-focused arguments can yield the kind of causal relevance that the emergentist is after. Until then, the emergentist will have to face up to the charges of epiphenomenalism.

Thus, the stage has been set. The emergentist must deal with problems of mental causation presented by the physicalist, and the physicalist must deal with problems of reducibility presented by the emergentist. The emergentist's problem is largely philosophical. Their task is to somehow account for mental efficacy, which seems to require a different conception that what was presented in this paper. The physicalist's problem is largely empirical. If we consider the historical development of emergentism, and its original claim that biological and chemical properties were irreducible emergent properties, it may seem the physicalist is simply waiting for the next scientific finding of some mechanism that allows for the reduction of phenomenal consciousness. The problem with this hope is that phenomenal consciousness is a *mystery*, and in many ways we don't even know where to begin. Until one of these problems is effectively dealt with, the status of emergence will be a truly open question, thus I suspect the debate surrounding emergence to be a lively one in philosophy of mind going forth.

Literature Cited

Alexander, S. (1966). *Space, Time, and Deity*. New York: Dover Publication.

Berent, E. (1995). Nonreducible Supervenient Causation. In E. Savellos and Ü. Yalçin (eds.), *Supervenience: New Essays*. Cambridge: Cambridge University Press. P. 169-186.

Bradden-Mitchell, D. (2007). Against ontologically emergent consciousness. In B. McLaughlin (ed.), *Contemporary Debates in Philosophy of Mind*. Oxford: Blackwell Publishing. P. 287-299.

Broad, C.D. (1925). *The Mind and its Place in Nature*. London: Routledge & Kegan Paul.

Bunge, M. (1977). Emergence and the Mind. *Neuroscience*, Vol. 2. Amsterdam: Elsevier. P. 501-509.

Chalmers, D. (1996). *The Conscious Mind*. Oxford: Oxford University Press.

-- (2006). Strong and weak emergence. *The reemergence of emergence: The Emergentist Hypothesis from Science to Religion*. Oxford: Oxford University Press. P 244-256.

Davidson, D. (1970). Mental Events. In L. Foster and J. Swanson(eds.), *Experience and Theory*. Amherst: University of Massachusetts.

Dretske, F. (1988). *Explaining Behavior: Reasons in a World of Causes*. Cambridge: MIT Press.

-- (1993). Reasons as Structuring Causes of Behaviour. In J. Heil and A. Mele (eds), *Mental Causation*. Oxford: Clarendon Press. P. 121-136.

Horgan, T. (2001). Causal compatibilism and the exclusion problem. *Theoria*, vol. 16. P. 95-116.

Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly*, vol. 32. Oxford: Blackwell Publishing. P. 127-136.

-- (1986). What Mary didn't know. *Journal of Philosophy*, vol. 83. Hanover, PA: Sheridan Press. P. 291-295.

Kim, J. (1989). Mechanism, purpose, and explanatory exclusion. *Philosophical Perspectives*, vol. 3. Oxford: Blackwell Publishing. P. 77-108.

-- (1990). Explanatory exclusion and the problem of mental causation. In E. Villaneuva (ed.),*Information, semantics and epistemology*. Oxford: Blackwell Publishing. P. 36-56.

-- (1993). *Supervenience and Mind: Selected philosophical essays*. Cambridge: Cambridge University Press.

-- (1997). Supervenience, emergence, and realization in the philosophy of mind. In M. Carrier (ed.), *Mindscapes: Philosophy, Science, and the Mind*. Pittsburgh: Pittsburgh University Press. P. 271-293.

-- (2000). *Mind in a physical world: An essay on the mind-body problem and mental causation*. Cambridge: MIT Press.

Literature Cited

-- (2006a). *Philosophy of Mind.* Cambridge: Westview Press.

-- (2006b). The Myth of Nonreductive Physicalism. In B. Beakley and P. Ludlow (eds.), *The Philosophy of Mind: Classic Problems/Contemporary Issues.* Cambridge: MIT Press. P. 427-442.

-- (2008). Making sense of emergence. In M. Bedau (ed.), *Emergence: Contemporary Readings in Philosophy and Science.* Cambridge: MIT Press. P. 127-154.

Lewis, D. (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy*, vol. 61. Victoria: Taylor and Francis Publishing. P. 343–377.

Lloyd Morgan, C. (1923). *Emergent evolution.* London: Williams and Norgate Publshing.

Loewer, B. (2007). Mental causation, or something near enough. In B. McLaughlin (ed.), *Contemporary Debates in Philosophy of Mind.* Oxford: Blackwell Publishing. P. 243-264.

Mill, J. S. (1843). *A System of Logic.* London: Longmans.

McLaughlin, B. (1992). The rise and fall of British emergentism. In M. Bedau (ed.), *Emergence: Contemporary Readings in Philosophy and Science.* Cambridge: MIT Press. 19-60.

-- (2008). Emergence and Supervenience. In M. Bedau (ed.),*Emergence: contemporary readings in philosophy and science.* Cambridge: MIT Press. P. 81-97.

Murphy, N. (2006). Emergence and Mental Causation. In P. Clayton and P. Davies (eds.), *The Re-Emergence of Emergence: The Emergentist Hypothesis from Science to Religion.* Oxford: Oxford University Press. P. 227-243

Nagel, E. (1961). *The Structures of Science.* London: Routledge & Kegan Paul.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, vol. 83. Durham, NC: Duke University Press. P. 435-450.

O'Connor, T. and Wong, H. (2012). Emergent Properties. In E. Zalta (ed.) *The Stanford Encyclopedia of Philosophy.* \url{http://plato.stanford.edu/archives/spr2012/entries/properties-emergent/

Pepper, S. C. (1926). Emergence. *The Journal of Philosophy*, vol. 23. New York: Columbia University Press. P. 241-245.

Popper, K. R., Eccles, J. C., John, C., & Carew, J. (1977). *The self and its brain* (Vol. 977, p. 362). Berlin: Springer International.

Putnam, H. (1975). The Meaning of 'Meaning.' *Minnesota Studies in Philosophy of Science*, vol. 7. Minneapolis: University of Minnesota Press. P. 131-193.

Sellars, R. W. (1922). *Evolutionary naturalism.* Chicago: Open Court Publishing Company.

Stoljar, D. (2009). Physicalism. In E. Zalta (ed.) *The Stanford Encyclopedia of Philosophy.* \url{http://plato.stanford.edu/archives/fall2009/entries/physicalism/

Literature Cited

Van Gulick, R. (2001). Reduction, Emergence and Other Recent Options on the Mind/Body Problem: A Philosophical Overview. *Journal of Consciousness Studies*, vol. 8. P. 1-34

Yablo, S. (1992). Mental Causation. *The Philosophical Review*. Durham, NC: Duke University Press. P. 245-280.

-- (1997). Wide causation. *Noûs*, vol. 31. Hoboken, NJ: Wiley-Blackwell Publishing. P. 251-281.