



AN EVOLUTIONARY ARGUMENT AGAINST PHYSICALISM

- or some advice to Jaegwon Kim and Alvin Plantinga



D-UPPSATS 15 HP, RELIGIONSFILOSOFI
Författare: Christoffer Skogholt
Handledare: Professor Mikael Stenmark
Opponent: Oliver Li
Datum för framläggning: 2014-09-18
VT 2014

Abstract

According to the dominant tradition in Christianity and many other religions, human beings are both knowers and actors: beings with conscious beliefs about the world who sometimes act intentionally guided by these beliefs. According to philosopher of mind Robert Cummins the “received view” among philosophers of mind is epiphenomenalism, according to which mental causation does not exist: neural events are the underlying causes of both behavior and belief which explains the correlation (not causation) between belief and behavior. Beliefs do not, in virtue of their semantic content, enter the causal chain leading to action, beliefs are always the endpoint of a causal chain. If that is true the theological anthropology of many religious traditions is false.

JP Moreland draws attention to two different ways of doing metaphysics: *serious* metaphysics and *shopping-list* metaphysics. The difference is that the former involves not only the attempt to *describe* the phenomena one encounter, it also involves the attempt of locating them, that is *explaining* how the phenomena is possible and came to be given the constraints of a certain worldview. For a physicalist these constraints include the atomic theory of matter and the theories of physical, chemical and biological evolution.

Mental properties are challenging phenomena to locate within a physicalist worldview, and some physicalists involved in “serious metaphysics” have therefore eliminated them from their worldview. Most however accept them, advocating “non-reductive physicalism” according to which mental properties supervene on physical processes. Even if one allow mental properties to supervene on physical processes, the problem of mental causation remains. If mental properties are irreducible to and therefore distinct from physical properties, as the non-reductive physicalists claim, they cannot exert causal powers if one accepts the causal closure of the physical domain – which one must, if one is a “serious physicalist” according to physicalist philosopher of mind Jaegwon Kim.

Alvin Plantinga, in his *Evolutionary Argument against Naturalism*, shows that if mental properties, such as the propositional content of beliefs, are causally inefficacious, then evolution has not been selecting cognitive faculties that are reliable, in the sense of being conducive to true beliefs. If the content of our beliefs does not affect our behavior, the content of our belief is irrelevant from an evolutionary standpoint, and so the content-producing part of our cognitive faculties are irrelevant from an evolutionary standpoint. The “reliability” – truth-conduciveness – of our cognitive faculties can therefore not be explained by evolution, and therefore not located within the physicalist worldview. The only way in which the reliability of our cognitive faculties can be located is if propositional content is relevant for behavior.

If we however eliminate or deny the reliability of our cognitive faculties, then we have abandoned any chance of making a rational case for our position, as that would presuppose the reliability that we are denying.

But if propositional content *is* causally efficacious, then that either – if we are non-reductive physicalists and mental properties are taken to be irreducible to physical properties – implies that the causal closure of the physical domain is false or - if we are reductive physicalists and not eliminativists regarding mental properties - it shows that matter *qua* matter can govern itself by rational argumentation, in which we have a pan-/localpsychistic view of matter. Either way, we have essentially abandoned physicalism in the process of locating the reliability of our cognitive faculties within a physicalist worldview. We have also affirmed the theological anthropology of Christianity, in so far as the capacity for knowledge and rational action is concerned.

Keywords: Philosophy of mind, mental causation, reductionism, physicalism, *the evolutionary argument against naturalism*, *the myth of nonreductive materialism*, Alvin Plantinga, Jaegwon Kim

Table of contents

Chapter 1. Introduction and background	3
1.1 Introduction.....	3
1.2 Purpose and method	5
1.3 Research overview and theoretical background.....	7
1.3.1 Reductionism - introduction.....	7
1.3.1.1 Causal reduction and ontological reduction – is there a difference?.....	8
1.3.1.2 Epistemological (or theoretical) reductionism	12
1.3.2 Philosophy of mind – an overview of the most important positions.....	14
Chapter 2. Physicalism, mentality and mental causation	22
2.1 Introduction.....	22
2.2 What do people (in general and Jaegwon Kim in particular) mean by physicalism?	23
2.3 Mental phenomena	27
2.4 Mental causation	31
2.5 The basic argument against non-reductive physicalism.....	32
2.6 Kim’s project of reducing mentality	33
Chapter 3. The Evolutionary Argument against Naturalism	39
3.1 Alvin Plantinga	39
3.2 The argument – introductory remarks.....	39
3.3 The argument – some conceptual clarifications.....	40
3.3.1 The expression “reliable” in reliable cognitive faculties.....	41
3.3.2 “Undercutting” and “rebutting” defeaters	45
3.4.1 The structure of the argument	45
3.4.2 Analyzing the argument	46
3.5 Why is this a problem?.....	54
3.6 Evaluation of the problem for the materialist.....	55
3.7 The proposed solution.....	55
3.8 Plantinga’s solution as inadequate for Plantinga (and other dualists of his kind).....	56
3.8.1 The proposed solution as inadequate for a what Plantinga calls “materialist”.....	58
Chapter 4. Conclusions and questions for further research	59
4.1 Summary and conclusions.....	59
4.2 Questions for further research.....	62

Chapter 1. Introduction and background

1.1 Introduction

In his book *Physicalism, or something near enough* (2005) Jaegwon Kim argues that the mind-body problem, for a physicalist like himself, consists of two problems:

1. *why is there, and how can there be, such a thing as the mind, or consciousness, in a physical world?*
2. *how can the mind exercise its causal powers in a causally closed physical world?*¹

These problems, the problem of consciousness and the problem of mental causation, are intimately connected with two capacities that are often taken to characterize us as human beings: we are *knowers* - beings who form conscious beliefs about reality - and we are *actors* - that is, we can act deliberately on the basis of our knowledge and not only unconsciously or reflexively. The difficulty of explaining how this might work in a physicalist ontology has led some physicalist philosophers to deny mental experience, and quite a few physicalist philosophers to deny mental causation.² The philosopher of mind Robert Cummins calls this latter position, known as epiphenomenalism, “the received view” and Jaegwon Kim notes that many brain scientists seem, at least implicitly, to have this view.³ This view is mostly motivated by the fact that we lack an adequate *explanans* - an explanation or account of how the mental might arise from the physical and/or of how mental causation *qua* mental might work; as a result such philosophers deny the existence of the *explanandum* - that which is in need of explanation, i.e. mental experience and/or mental causation. It is generally agreed that rejecting mental experience is self-referentially incoherent; the purpose of this essay is to argue that it is also implicitly self-referentially incoherent to deny mental causation, if we believe that our capacity for having a mental life has evolved through the process of evolution.

Anyone reading this sentence is, in some sense, an observer. That means that, at the very least, you are conscious and you have mental experiences. Unless you are a solipsist and believe that any experience, including the experience of reading this essay, is produced entirely by your own mental life, the reading of this essay also means that you can have conscious experiences of something which exists independently of yourself.⁴ So, then you would be an observer, not only of your own mental life but also of other things, which exist before you are aware of them, before they enter into your mental life (such as this essay). In any case, there are observers: subjects that have conscious experiences.

¹ Kim, Jaegwon *Physicalism, or something near enough* (2005) p. 13

² Those denying mental experience are called eliminative materialists, see: Ramsey, William “Eliminative Materialism” at Stanford Encyclopedia of Philosophy <http://plato.stanford.edu/entries/materialism-eliminative/> (retrieved: 2014-06-18)

³ Robert Cummins was quoted by Plantinga in Beilby, James (ed.) *Naturalism Defeated? Essays Plantinga's Evolutionary Argument against Naturalism* (2002) p. 7 Jaegwon Kim writes about the implicit epiphenomenalism of brain scientists in his *Philosophy of Mind* (2011) p. 97

⁴ “solipsism” in *Nationalencyklopedin* <http://www.ne.se/lang/solipsism> retrieved: 2014-08-02.

Anyone denying that is self-referentially incoherent - the person would presumably have the experience that she denied having any experiences.

But are there also *actors* in the world? There are certainly *events* (such as conscious experiences). Apart from reading sentences I have also had a few other experiences, such as the experience that it rains and that people walk and that people talk. What causes these events or these experiences? If we are not solipsists some of these events are caused (if they are caused at all) by something outside ourselves, and sometimes they are (seemingly at least) caused by other people. Sometimes people can be causes of events involuntarily (such as when they are reacting reflexively) and sometimes they act voluntarily, intentionally. In sociology the distinction between behavior and action is made in that way: behavior is automatic, action is carried out voluntarily and intentionally, so that an act is done with specific conscious goals.⁵ The difference between behavior and action is that actions are done in virtue of intentions. Intentions are beliefs and desires, phenomena which are part of the mental life of observers.⁶

Can intentions and desires enter into the causal chain leading to action? Can they do so *as* intentions and desires? That is, regarding beliefs, in virtue of their propositional content and, regarding desires, in virtue of their felt or phenomenal aspects? I believe they can, and the purpose of this essay is to argue that if they cannot, then we have no reason to believe any proposition, including the proposition that propositional content isn't causally relevant for action. This has to do with the fact that the propositional content of our beliefs would then be irrelevant for our behavior, and evolution is, "interested (so to speak) only in adaptive behavior".⁷ That is, if it is evolution by natural selection that is responsible for our belief-producing faculties (or our cognitive faculties as they are more commonly called) then, if the propositional content of our beliefs is irrelevant for our behavior, evolution has not selected belief-producing faculties that would be conducive to the formation of true rather than false beliefs. The content, and *a fortiori*, the truth or falsehood of our beliefs would be utterly irrelevant from a behavioral and therefore evolutionary standpoint. Alvin Plantinga quotes the materialist philosopher Patricia Churchland who drives home this point rather forcibly:

Looked at from an evolutionary point of view, the principal function of nervous systems is to enable the organism to move appropriately. Boiled down to essentials, a nervous system enables the organism to succeed in the four F's: feeding, fleeing, fighting, and reproducing. The principal chore of nervous systems is to get the body parts where they should be in order that the organism may survive. (...) Improvements in sensorimotor control confer an evolutionary advantage: a fancier style of representing is advantageous *so long as it is geared to the organism's way of life and enhances the organism's chances of survival*. Truth, whatever that is, definitely takes the hindmost.⁸

⁵ Engdahl, Oskar; Larsson, Bengt *Sociologiska perspektiv: grundläggande teorier och begrepp* (2006) p. 31

⁶ See for instance Siewert, Charles "Consciousness and Intentionality" in *Stanford Encyclopedia of Philosophy* (<http://plato.stanford.edu/entries/consciousness-intentionality/>) Retrieved: 2014-08-25

⁷ Plantinga, Alvin *The Evolutionary Argument against Naturalism: An Initial Statement of the Argument* in Beilby, James (ed.) *Naturalism Defeated?* (2002) p. 4

⁸ Patricia Churchland was quoted in Plantinga, Alvin *Warrant and Proper Function* (1993) p. 218 (Churchland's emphasis)

I will argue for my position by analyzing Alvin Plantinga's famous *Evolutionary Argument against Naturalism* and I will try to show that this argument can be put to better use as an argument against reductionism in philosophy of mind, which I will define as the position that the mental cannot exert causal influence as mental (in virtue of the propositional content of beliefs or the phenomenal or felt aspects of desires) but only in virtue of its neurophysiological properties.

Seen from a commonsense viewpoint, the question of mental causation might seem trivial, as a matter of course. But in philosophy, where beliefs are usually put in explicit relationship to each other and evaluated together, mental causation is highly controversial. This is mainly because the default metaphysical position for many philosophers and for many in the western culture in general is (some version of) physicalism, according to which everything that exists is physical.⁹ And it is, as Plantinga says:

extremely hard, given materialism, to envisage a way in which the content of a belief could get causally involved in behavior. If a belief just is a neural structure of some kind – a structure that somehow possesses content – then it is exceedingly hard to see how content can get involved in the causal chain leading to behavior.¹⁰

Hard, yes, but necessary if one wants to affirm the reliability of our cognitive faculties. If that is true, the question arises what form of materialism that is possible, if one accepts mental causation.

How is this relevant for philosophy of religion? It is relevant because most religious worldviews (and many secular worldviews and several, if not all, academic disciplines within the humanities and the social sciences, for that matter) include or imply the claim that mental phenomena can be causes of behavior.

1.2 Purpose and method

1. The major purpose of this essay is to argue against reductionism in philosophy of mind, which I believe is a philosophical perspective that is often implicitly or explicitly understood as being supported by science, and which cannot be accommodated within a religious worldview, at least not with the mainstream traditions of Christianity. I take reductionism in philosophy of mind to mean that the mental cannot exert causal power in virtue of its mental properties. I will argue against this position by a modification of Alvin Plantinga's *Evolutionary Argument against Naturalism* (EAAN).

In an extended way, then, my essay can be seen as belonging to that field of reflection within philosophy of religion and philosophical theology known as the science and religion-debate. In so far as reductionism is claimed to be supported by science and in so far as it is in conflict with the anthropologies of most religions, this essay belongs to that field of reflection. It is furthermore my thesis that in order to escape reductionism in philosophy of mind, you must abandon physicalism,

⁹ Stoljar, Daniel "Physicalism" *Stanford Encyclopedia of Philosophy*
<http://plato.stanford.edu/entries/physicalism/> (retrieved: 2014-05-20)

¹⁰ Plantinga, Alvin in Beilby, James (ed.) *Naturalism defeated?* (2002) p. 10

as that position is traditionally understood, which is of course also often taken to be a philosophical position supported by science.

2. One of the most influential physicalist philosophers of mind today is Jaegwon Kim. I will present and analyze his argument against non-reductive materialism, which I believe is sound, and his proposed solution, which I do not think is a (working) solution to the problem of mental causation. The reason that I present and discuss his argument is that it is helpful for illustrating where the problem, for physicalism, lies with respect to mental causation.
3. Apart from arguing against physicalism I also want to show that the by Plantinga implied solution to the problem stated in EAAN is inadequate both for a dualist like Plantinga and for a Christian who is, what Plantinga calls, a “materialist”. Some clarification is needed here. Whereas for Kim anyone is a dualist who is not a reductive materialist, for Plantinga anyone is a materialist who is not a substance dualist. That means that someone like Philip Clayton who advocates strong emergence and property dualism is a dualist in Kim’s view, and a materialist in Plantinga’s view. The reason is that for Plantinga supervenience and emergence are materialist positions, since they claim that although mentality is irreducible to physical properties, mental properties arise from physical structures and processes. Plantinga thinks that that is impossible, and for that reason (among others) he argues for substance dualism, and an immaterial soul which has the ability to think among its basic properties.¹¹

For Kim, however, anyone who isn’t a reductive physicalist or an eliminativist is in the end a dualist. So therefore he argues that the only positions that exists are: dualism, reductionism and eliminativism and that no form of dualism (including property dualism) is admissible from a physicalist point of view.¹²
4. The conclusion of my essay is that in order for us to be justified in trusting our cognitive faculties, given that we believe that they have been produced by the process of evolution, we must affirm mental causation. Affirming mental causation *qua* mental means that the resulting anthropology is compatible with a theological anthropology, at least as far as the capacity for knowledge and rational, intelligent action is concerned, which is not the case with the reductionist position which denies mental causation. Since denial of mental causation is quite common, even “the received view”¹³ according to some philosophers of mind, I argue that this is a non-trivial conclusion.

¹¹ Plantinga, Alvin “Materialism and Christian Belief” in van Inwagen, Peter; Zimmerman, Dean (ed) *Persons – Human and Divine* (2007)

¹² Kim, Jaegwon “The Myth of Nonreductive Materialism” (1989) <http://www.jstor.org/stable/3130081> retrieved: 2014-05-01

¹³ Plantinga, Alvin in Beilby, James (ed.) *Naturalism Defeated?* (2002) p. 6 (Plantinga quotes Robert Cummins who calls it the received view)

Another non-trivial conclusion is that affirming mental causation also has implications for our ontology in general: if mental causation *qua* mental is real, that implies either, if mental properties are irreducible to physical properties, that the principle of the causal closure of the physical domain is false, or if we understand mental properties as physical that our understanding of the nature of the physical must be radically expanded, as to include mental properties *qua* mental - and allowing these properties to have causal powers, so that it is sometimes true that “an event, *in virtue of its mental property*, causes another event to have a certain physical property” as Jaegwon Kim puts it.¹⁴

1.3 Research overview and theoretical background

In a sense, the central problem of this essay is the question of mental causation (can the mind exert causal influence *qua* mental?). The problem of mental causation ranges over several different fields within philosophy and science, and it is obviously not possible to give an overview over all areas that are of relevance for this issue, but some of the more fundamental themes in philosophy of mind can be mentioned. Since the modern debate in philosophy of mind is closely connected to the issue of reductionism, I will include a section on reductionism before presenting the different positions in modern philosophy of mind, so as to be able to understand them in relationship to the different forms of reduction.

1.3.1 Reductionism - introduction

Reductionism is often said to be of at least three different kinds: ontological, epistemological and methodological.¹⁵ Anne Runehov argues that all forms (she lists five) of reductionism can be reduced to two: ontological and methodological.¹⁶ John Searle argues that “ontological reduction” is what all the others are aiming at – and I believe that is true of the different forms of epistemological reduction, but not of methodological reduction which is just a form of self-limitation of one’s study without any ontological commitments.¹⁷

Searle further writes that the “basic intuition that underlies the concept of reductionism seems to be that certain things might be shown to be *nothing but* certain other sorts of things.”¹⁸ Jaegwon Kim seems to agree when he writes that the “root meaning” of reductionism was given by the materialist philosopher of mind John JC Smart when he said that “sensations are nothing ‘over and above’ brain processes”.¹⁹ Jaegwon Kim's understanding of reduction as ontological reduction is furthermore indicated by the fact that he writes that “if anything is physically reduced, it must be identical with

¹⁴ Kim, Jaegwon “The Myth of Nonreductive Materialism” (1989) p. 43 Italics in original

¹⁵ Barbour, Ian *Religion in an Age of Science* Volume 1 (1990) p. 165-168

¹⁶ Runehov, Anne *Sacred or Neural? Neuroscientific Explanations of Religious Experience: A Philosophical Evaluation* (2004) s. 232

¹⁷ Although it sometimes slips over into a form of implicit ontology, as sometimes is the case between methodological and ontological naturalism.

¹⁸ Searle, John R. *The Rediscovery of Mind* (1994) p. 112 italics in original

¹⁹ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 34

some physical item”.²⁰ That is, you show that something which seems to be something “over and above” something else really is not: “If Xs are reduced to Ys, then Xs are nothing over and above Ys.”²¹ So when Kim advocates reduction of mental properties to physical properties, it is a form of ontological reduction: the identification of the mental with the physical. Runehov likewise describes ontological reductionism as the position that some “phenomena, processes or events can be exhaustively described by other phenomena, processes or events”.²² So the paradigmatic case of reduction is the ontological reduction and ontological reduction is the process in which you identify something with something else.

As a preliminary reflection I would like to say that, so understood, it seems as if an ontological reduction must either result in an expansion in the understanding of the “Ys” (that which the Xs are being reduced to) or it will result in an elimination of at least some of the properties that are taken to characterize Xs (that which is being reduced). The situation is that *phenomenologically speaking* Xs seems to have characteristics that Ys do not – otherwise there would be no reduction of Xs in the identification of Xs with Ys. And either our reduction will involve some kind of elimination of these characteristics, or they will be shown to be part of the properties of Ys. Since an identification is the symmetrical identification of something with something else, it must either involve the elimination of some properties of the phenomenon to be reduced, or an expansion of that to which it is supposed to be identified, which did not seem to have those properties. When the physicalist Galen Strawson argues that physicalists should be panpsychists, that is a form of identification in which the “Ys” get properties thought to belong exclusively to “Xs”.²³ When the eliminativists Paul and Patricia Churchland argue that physicalists ought to be eliminativists, that is an argument that has the form that since Xs are nothing over and above Ys and Ys are not conscious, neither are Xs.

1.3.1.1 Causal reduction and ontological reduction – is there a difference?

The first problem of ontological reduction is if there is a difference between ontologically reducing X to Y, and giving a causal account of X in terms of Y.²⁴ Searle discusses the relationship between causal and ontological reduction in chapter five in *The Rediscovery of the Mind*. To understand this discussion we need to make a distinction between properties which belong to an object, let us call

²⁰ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 34

²¹ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 34

²² Runehov, Anne *Sacred or Neural? Neuroscientific Explanations of Religious Experience: A Philosophical Evaluation* (2004) p. 233

²³ Strawson, Galen “Realistic Monism: Why Physicalism Entails Panpsychism” *Journal of Consciousness Studies* Vol. 13, No. 10-11 (2006) available at:

http://www.utoronto.ca/~seager/strawson_on_panpsychism.doc (Retrieved: 2014-08-01)

²⁴ If it is true that “wholes” may be more than the sum of their parts, it seems as if pointing out the causes of x isn’t enough for describing x. Of course this does not mean that some (perhaps many) objects indeed are possible to exhaustively define by pointing to their causes.

them *basic properties*, and properties which belong to “a system” – that is a collection of objects – let us call them (as Searle does) *system features*.²⁵

Now there are two forms of system features that Searle discusses, one of them can be seen as a sort of aggregated effect of the individual objects in the system, such as mass and shape.²⁶ The other kind – “causally emergent system features” – are not merely the aggregated effect of the individual parts, but must be accounted for by reference to the interaction between these parts. Here temperature, solidity and liquidity are taken as examples.²⁷

These system features are objective features of the world, i.e. they have what Searle calls “observer independent effects” although they can be given a causally reductive account in terms of the behavior of the parts of the system: the state of matter can be explained by reference to the movement of molecules. Searle also calls these system features “surface features” – to distinguish them from their underlying causes. Searle also thinks of subjective experiences as “surface features” having underlying causes (such as brain processes). So both physical “system features” and “subjective experiences” are called “surface features”. A causal reduction is a reduction in which you describe what causes these surface features, and the ontological reduction consists in *identifying* these surface features with their underlying (microphysical) causes.

Now, in the case of phenomena like heat or sound that implicitly involve both a notion of an object (that which causes the sound) and a conscious experience (the hearing of the sound) the causal reduction can only lead to an ontological reduction if you also redefine the term “heat” or “sound” to mean: the physical stimuli which the senses in question (vision and hearing) was stimulated by:

“We did not really eliminate the subjectivity of red, for example, when we reduced red to light reflectances; we simply stopped calling the subjective part “red”.”²⁸

That is, it is impossible to *define* something which involves a subjective experience in terms of non-conscious processes – although they may very well be the causes - unless you first change the meaning of the conscious experience so that it can be identified with these unconscious causes. So when we now in this scientific day and age define the color red as a certain wavelength, it is not the experience of seeing red we are talking about, it is the physical stimulus which, in certain circumstances and attended to by a conscious organism, will give rise to that experience.²⁹

²⁵ It is an interesting question what happens with this distinction if individual objects can combine to form genuine wholes. It seems as if most people consider atoms to be objects, but atoms are made up by smaller objects, elementary particles. Are the properties of atoms basic properties or system features? The same is true of molecules, and cells, and so on. Either one would have to say that the only “basic properties” are those that the fundamental units of matter have (elementary particles) and all other properties are “system features” – or one would have to acknowledge that in principle objects can combine to build new objects, with new “basic properties”.

²⁶ Searle, John *The Rediscovery of the Mind* (1992) p. 111

²⁷ Searle, John *The Rediscovery of the Mind* (1992) p. 111

²⁸ Searle, John *The Rediscovery of the Mind* (1992) p. 123

²⁹ The stimulus in itself is never enough for an experience of course. There must also be conscious beings processing the physical stimulus. Therefore we are only describing one of the causes of the experience, when we talk about wavelengths.

When we encounter the phenomenon of consciousness, the defining feature is that it is a subjective, first-person perspective and encounter with the world. This experience may, or may not, be caused by certain brain processes. But this experience cannot be reductively identified with these brain processes, *as these are described in the natural sciences that study them*. Why? For the simple reason that conscious experience is not a property that occurs in neuroscience. The only way in which we can go from causal reduction to ontological reduction in this case is through a redefinition of the term consciousness.³⁰ The only way in which a causal reduction could lead to an ontological reduction to these causes, that is an identification with these causes, would be if we allowed the causes themselves to be conscious – or through an elimination of the conscious element in the phenomenon to be reduced.

The redefinition of “heat” and “color” was done by excluding the subjective aspect from the definition, so the description is now only a description of the object that causes the experience. These subjective aspects are often called “secondary qualities” since they are observer relative, and are distinguished from “primary qualities” which are often taken to be objective, observer-independent features of the world. “Seeing red” is a secondary quality, the wave-length of light is a primary quality.

According to Searle, conscious experiences cannot be identified with their nonconscious causes, so causal reduction is not enough for ontological reduction in the case of consciousness. The only way to accomplish that identification would be to redefine consciousness so as to exclude its conscious aspects. What about other “causally emergent system features” – are they also irreducible to their underlying causes? It would perhaps seem that “surface features” which are primary qualities and objective features of the world ought to be an irreducible part of the furniture of the universe. Searle is a bit unclear, though, if you can make an “ontological reduction” – that is the *identification of* surface features of the non-conscious kinds to their underlying causes - even when these surface features are taken to be objective and have observer-independent effects. He acknowledges that this involves a certain redefinition, but does not seem to object to it:

“In general, in the history of science, successful causal reductions tend to lead to ontological reductions. Because where we have a successful causal reduction, we simply redefine the expression that denotes the reduced phenomena in such a way that the phenomena in question can now be identified with their causes.”³¹

As for the surface features in the forms of conscious experience, this took place by changing the meaning of the term so that it no longer referred to subjective experiences, but only the object. But what about the unconscious, objectively existing, system-features that have underlying causes, can they be identified to these causes?

“Solidity is defined in terms of the vibratory movements of molecules in lattice structures, and objective, observer-independent features, such as impenetrability by other objects, are now seen as surface effects of

³⁰ Searle, John *The Rediscovery of the Mind* (1992) p. 117

³¹ Searle, John *The Rediscovery of the Mind* (1992) p. 115

the underlying reality. Such redefinitions are achieved by way of carving off all of the surface features of the phenomenon, whether subjective or objective, and treating them as effects of the real thing.”³²

Searle thinks that impenetrability is an objective feature of the world, it is a “primary quality” which has an underlying (microphysical) cause, but it may be identified with its microphysical cause. It seems as if Jaegwon Kim argues that such a causal reduction is not necessarily enough for a reductive identification:

“This baseball has causal powers that none of its proper parts, in particular none of its constituent microparticles, have, and in virtue of its mass and hardness it can break a window when it strikes it with a certain velocity. The shattering of the glass was caused by the baseball and certainly not by the individual particles composing it.”³³

Kim would of course agree with Searle that the surface feature of “hardness” is caused by these underlying causes - still it seems as if Kim does not *identify* this feature with those underlying causes. Kim seems here to affirm a form of causal power at the macrolevel – the baseball - and in order to do so, the baseball must be understood as existing as an object in its own right. Kim writes this in a context in which he wants to make the case that his view of supervenience does not exclude causality at levels above the microphysical. Unless this is just a rhetorical strategy to escape certain unpleasant consequences of his supervenience argument and we take him at face value, it clearly implies that the baseball is not identical to its constituent parts. Perhaps a baseball, though, is a rather straightforward example of an object that can be reductively identified with its underlying causes. The important thing to note is that this is not *necessarily* the case. For Searle, conscious experience is not identifiable with its causes - and possibly that is not always the case even for physical objects.

Furthermore, regarding physical phenomena, it is evident that the *structure*, which is a property of the whole, can be relevant for understanding the nature of an object or system. This is perhaps most easily understood with reference to objects or systems that perform certain functions, say a windmill. This structure of course also has underlying microphysical causes (or realizers), but the windmill is not identical to them. Functions are realizable in a multiple ways, that is, they can be realized by different microphysical causes, which serves to show that they are not identifiable with these material causes. Their nature is defined abstractly, by the function they perform rather than by reference to the physical particles that they are composed of.³⁴

This is evident also when we consider objects intended to express meaning. A painting has a motive which is only visible at the “surface level” although it is caused by a lot of individual pixels (the underlying reality). A sentence has a meaning at the level of the “whole” and cannot be understood only as the aggregated effect of individual words. In these examples these surface features cannot be identified with their “underlying” or microstructural causes. In short: it does not seem as if a phenomenon – conscious or not - is necessarily identical to its underlying causes.

³² Searle, John *The Rediscovery of the Mind* (1992) pp. 119

³³ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 56

³⁴ Maslin, Keith T. *An Introduction to the Philosophy of Mind* (2007) p. 133

1.3.1.2 Epistemological (or theoretical) reductionism

In the debate about reductionism the center of discussion is often whether the “special sciences” (all sciences apart from physics) are reducible to the “basic science” (physics). As Alyssa Ney writes:

The type of reductionism that is currently of most interest in metaphysics and philosophy of mind involves the claim that all sciences are reducible to physics.³⁵

The philosopher Sven-Ove Hansson points out in an article about science and reductionism that although the special sciences could in principle be reducible to physics, they might not be so in practice, because of our lack of knowledge.³⁶ It might be totally incalculable to explain the behavior of a family using only the fundamental laws of physics. So for practical reasons we use other concepts, which in principle, it is argued, could be reduced to concepts within physics. If such an explanatory reductionism were to take place, then that would be an example of epistemological reductionism. Epistemological reductionism can be seen as the actual attempt to show that the claim made by the ontological reductionist is true. Sven-Ove Hansson writes that the epistemological reductions that have been in fact carried out are “extremely modest” in comparison to the claim made by the ontological reductionist. The important thing to see is that the possibility of ontological reduction implies the *reducibility* in principle, in the epistemological sense, but not that it has been achieved in practice. I used to think that if an epistemological reduction was successful, then that would be enough for ontological reduction. However, if epistemological reduction is understood as finding the underlying causes of a “system feature” as it sometimes is, then epistemological reduction is not necessarily enough for ontological reduction.

Alyssa Ney lists three forms of epistemological (or inter-theoretic) reduction: reduction as translation, reduction as derivation and reductionism as explanation.³⁷ The first version is associated with the logical positivists Rudolf Carnap and Otto Neurath, the second with the logical positivist Ernest Nagel and the third with John Kemeny and Paul Oppenheim. There is considerable overlap between these forms of reductionism.

Reduction as “translation” means that concepts from one special science (such as the concept “emotions” in psychology) can be translated, without loss of meaning (it is alleged) to concepts within a universal scientific language, something Carnap called “physical language, understood as the language of objects in space and time (rather than physics per se)”.³⁸ The claim is that nothing in the concept “emotions” that has a real application (refers to something real) is impossible to capture in this physical language. Physical language is able to capture everything that the concept emotion refers to,

³⁵ Ney, Alyssa “Reductionism” in *Internet Encyclopedia of Philosophy* <http://www.iep.utm.edu/red-ism/> retrieved: 2014-05-13

³⁶ Hansson, Sven-Ove “Vetenskap och reduktionism” <http://www.vof.se/folkvett/ar-1991/nr-1/vetenskap-och-reduktionism/>

³⁷ Ney, Alyssa “Reductionism” in *Internet Encyclopedia of Philosophy* <http://www.iep.utm.edu/red-ism/> retrieved: 2014-05-13

³⁸ Ney, Alyssa Reductionism in *Internet Encyclopedia of Philosophy* <http://www.iep.utm.edu/red-ism/> retrieved: 2014-05-13

and which actually exists. Since language about objects in space and time is translatable to physical theory, in the end it seems that all of science should be reducible to physics, if all science is translatable to language about objects in space and time.

One motivation for this kind of reductionism is the goal of unity in science, and that one should not explain the same type of phenomenon with several different concepts, if it can be adequately explained and described by one type of concept.³⁹ Neurath worried that we could get a fragmented science and also logical conflicts between scientific disciplines if we had different irreducible concepts in different sciences. Neurath worked from the premise that the world is not fragmented, but a unity, and science is the attempt to describe this world.⁴⁰ So different disciplines should not conflict each other, and all of their findings should be describable in one and the same type of language. When saying that this language should be a “physical language” it does seem to imply an *a priori* restriction to the phenomena that can be real, since everything that is real is presumed to be physical.

Ernest Nagel describes epistemological reductionism as the aim to show how the laws described in a special science can be shown to be logical consequences of the assumptions of a more basic science. Nagelian reduction is for that reason often called “nomological reduction” and it is also nomological *derivation*: you derive the laws or causal processes described within a special science from what you know from the base science. Nagel distinguished between “homogeneous” and “heterogeneous” reduction. The former implies that the concepts in a special science can be translated into concepts in the base science. The latter, which Nagel thought was the more common one, does not involve such a claim. Instead there is a reduction using bridge laws, connecting terms in different sciences.

Nagel argued that a successful reduction consists in deriving from a base science a process which a special science describes, using concepts in the base science, connecting them to events in the special science, using “bridge laws”. So for instance, let us say that a process described in psychology is that when you understand that there is a threat, you feel fear. Let “S1” be the event in the special science that you understand that you face harm and “S2” be the phenomenon, described in the special science, that you feel fear. There is now some kind of neurological event B1 that creates S1 and some neurological event B2 that creates S2.

However if we did not have (from neuroscience) any independent knowledge of S1 and S2, we would not out of what we currently know about neurons or the neurological events B1 and B2 be able to derive the existence of S1 and S2. We therefore need to postulate “bridge laws” which states that B1 and B2 gives rise to S1 and S2.⁴¹ However as of today those bridge laws are unknown. But let us

³⁹ Ney, Alyssa “Reductionism” in *Internet Encyclopedia of Philosophy* <http://www.iep.utm.edu/red-ism/> retrieved: 2014-05-13

⁴⁰ Ney, Alyssa “Reductionism” in *Internet Encyclopedia of Philosophy* <http://www.iep.utm.edu/red-ism/> retrieved: 2014-05-13

⁴¹ Batterman, Robert “Intertheory Relations in Physics” in *Stanford Encyclopedia of Philosophy* <http://plato.stanford.edu/entries/physics-interrelate/> (retrieved: 2014-05-14)

say that they exist. What that implies is that the knowledge of B1 and B2 is incomplete unless we also include that (and how) B1 and B2 gives rise to S1 and S2.

In what sense is this a reduction? For me, it seems more to be an expansion of the base science, rather than a reduction of the special science, to be able to include that which the special science talks about. We could say that S1 and S2 are epiphenomena, that is, they are unable to exert causal powers (B1 gives rise to S1 and B1 gives rise to B2 which in its turn gives rise to S2). But S1 and S2 still exist, and they are not the same as the current neuroscientific description of B1 and B2. The bridge-laws then seem to be almost magical devices that do not explain anything, they just point out the fact that these events are correlated. If we say that B1 is such that S1 follows, but S1 is not eliminated, this might imply an expansion in the understanding of B1.⁴²

Carl Hempel had a similar view to Nagel of reduction as including derivation, but also including a *sort of* translation of the concepts to be reduced. I include this quote, because it is illuminating for understanding the position in philosophy of mind known as the mind/brain identity-theory:

Carl Hempel saw the reduction of a theory as involving two tasks. First, one reduces all of the *terms* of that theory, which involves translation into a base language. As Hempel notes, “the definitions in question could hardly be expected to be analytic... but ... may be understood in a less stringent sense, which does not require that the definiens have the same meaning, or intension, as the definiendum, but only that it have the same extension or application” (1966, 103). Then, one reduces the *laws* of the theory into those of a base theory by derivation (1966, 104).⁴³

Reduction as explanation, lastly, means that you show that a set of observations which are explainable in a special science can also be explained in the base science, in which case the special science theory is made redundant and superfluous. The difference seems to be that you are not trying to reduce laws, but individual events. If all causality is “nomological” (law-based) as Donald Davidson argues, reduction as explanation seems to presuppose reduction as derivation, and then be the application of that reduction to a specific set of observations.

1.3.2 Philosophy of mind – an overview of the most important positions

Jaegwon Kim says that René Descartes (1596-1650) “invented” the mind-body problem.⁴⁴ However, Karl Popper objects to that popular description pointing out that the issue has been around at least since the beginning of recorded philosophy.⁴⁵ What Descartes did however was to situate the problem in the context of the scientific understanding of matter of his time, the mechanistic worldview and thereby giving the mind-body problem its modern flavor: how shall we understand the mind in relationship to what we know about the world through the sciences? Descartes' solution was to affirm a dualism of substances, material (*res extensa*) and mental (*res cogitans*) and thereby excluding the

⁴² As it turns out, this seems to be the criticism by Jaegwon Kim against Nagelian reduction. (See below.)

⁴³ Ney, Alyssa “Reductionism” in *Internet Encyclopedia of Philosophy* <http://www.iep.utm.edu/red-ism/> (retrieved: 2014-08-01)

⁴⁴ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 8

⁴⁵ Popper, Karl; Eccles, John *The Self and Its Brain* (1977)

mental sphere from the sphere where scientific research was made. This dualism made it more understandable how there could be mental properties, but had the price that it became harder to understand the interaction between these two seemingly disparate ontological categories. The different positions within philosophy of mind can be understood in relation to how they respond to the questions: is it a part of the natural order, is the mind physically reducible and is it causally efficacious *qua* mental? In this section I will give an overview of the most important positions in philosophy of mind during the 20th and 21st centuries.

Since ontological reduction is the “root meaning” of reductionism and ontological reduction means that two phenomena that seem to be distinct from one another are in reality one and the same, *reductive materialism* means that the mind and the brain are identical, or at least that the mind is identical to something physical.

There are a few different ways of understanding what this means, and at one extreme we have eliminative materialism which simply denies the existence of conscious, mental states: “beliefs, hopes, fears, desires, etc. do not really exist”⁴⁶ and the analytical behaviorists who would say that mental states do exist (they have extension) but their meaning, their “intension” is actually not conscious mental states, it refers to “possible and actual behavior”.⁴⁷ To refer to the different forms of reduction, analytical behaviorism is a form of “reduction as translation”: it is the claim that statements about subjective, conscious mental states can be translated, without loss of meaning, to statements about actual or potential behavior: “The mind does not cause the behavior: it *is* the behavior.”⁴⁸ But since this leaves out the first-person perspective in the definition of consciousness, it is an elimination of that with which we began our project of reduction.

The eliminative materialist form of reduction is not so much a reduction as it is a denial of there being anything – even phenomenological speaking - in need of reduction.⁴⁹ An interesting aspect of the relation between phenomenology and ontology when it comes to consciousness is that it does not seem to apply. It does not make sense to say that it *seems* to a subject that she has a conscious experience while in reality she hasn’t. As Roger Scruton puts it: “In the subjective sphere being and seeming collapse into each other. In the objective sphere they diverge.”⁵⁰ So, I would say, if eliminativism isn’t true, the prospect of ontological reduction of mentality is quite hard, because reducing in general means making the claim that how things seem is not what they really are.

The next form of reductive materialism is the so-called mind/brain identity theory. Identity theorists usually distinguish their position from the analytical behaviorists by pointing out that talk

⁴⁶ Searle, John *The Rediscovery of the Mind* (1992) p. 47

⁴⁷ Maslin, Keith T *An Introduction to the Philosophy of Mind* (2007) p. 101

⁴⁸ Maslin, Keith T *An Introduction to the Philosophy of Mind* (2007) p. 99

⁴⁹ Most people are weary that they are being unfair to the position, when they present it, as it seems so extreme. I share this worry, but I note that I am not alone. See Ramsey, William “Eliminative Materialism” in *Stanford Encyclopedia of Philosophy* <http://plato.stanford.edu/entries/materialism-eliminative/> retrieved: 2011-08-20 and Searle, John *The Rediscovery of the Mind* (1992) p. 6

⁵⁰ Scruton, Roger *Kant – A Very Short Introduction* (2001) p. 19

about mental states is not translatable without loss of meaning to talk about brain states.⁵¹ They deny that the expression “I have a pain in my back” has the same meaning or “intension” as “My c-fibers are firing” but argue that these expressions have the same “extension” or referent: “‘Sensation’ and ‘brain process’ may differ in meaning and yet have the same reference.”⁵²

This is usually exemplified by the expressions “the evening star” and “the morning star” which may be understood as being different in meaning: one refers to a phenomenon in the sky that appears in the evening and has certain characteristics, and another refers to a phenomenon that appears in the morning. They are different in meaning, but as we now know, both phenomena are caused by one and the same heavenly body, namely Venus, which then is the referent of both expressions although they are different in meaning. Thomas Aquinas argues in this way when he says that everyone seeks God, when they seek happiness, although the expressions “I am seeking happiness” and “I am seeking God” are not the same in meaning. We could also exemplify it by talk about Dr. Jekyll and Mr. Hyde which are clearly different in meaning (since Dr. Jekyll is a mean character and Mr. Hyde a nice one) but do refer to one and the same person. There is a problem here, though, and that is that the examples are disanalogous to the claims that the mind/brain identity theorists actually make regarding the mind and the brain. When they say that “the mind is the brain” they are actually not saying that there exists something, let us call it x, which has both types of properties, as is the case (mutatis mutandis) with Venus or my example with Dr. Jekyll and Mr. Hyde. They are ultimately claiming that only those processes and properties that go under the name “brain processes” exists. It is therefore more like claiming that there is no heavenly body which has the properties of the morning star, only a heavenly body which has the properties of the evening star.

Arguing like this, is not in actual practice arguing for the identity of the mind with the brain: it is denying the mind and affirming the brain instead. Identities are symmetrical, which means that if x is identical to y, then also is y identical to x.⁵³ If the identity-claims were taken seriously by identity theorists it would lead them to a panpsychist position in which matter would be understood to have among its basic, irreducible properties mental properties. But the mind/brain identity theorists argue instead as an idealist would, if she would say: since the mind and the brain are identical, the brain is identical with mental states and then go on denying that there was something physical at all, since mental states are nonphysical. The mind/brain identity theorist argues like this, but in the opposite direction, as Keith T. Maslin points out:

“What it does, in effect, is to get rid of mental features in favor of brain features. It claims, effectively that the end of the day, there are only brain features.”⁵⁴

⁵¹ Smart, JJC “The Mind/Brain Identity Theory” in *Stanford Encyclopedia of Philosophy* (<http://plato.stanford.edu/entries/mind-identity> retrieved: 2014-08-30)

⁵² Smart, JJC “The Mind/Brain Identity Theory” in *Stanford Encyclopedia of Philosophy* (<http://plato.stanford.edu/entries/mind-identity> (retrieved: 2014-08-30) The words *intension* means meaning and *extension* means referent. The word “unicorn” has an intension but no extension.

⁵³ Maslin, Keith T *An Introduction to the Philosophy of Mind* (2007) p. 76

⁵⁴ Maslin, Keith T *An Introduction to the Philosophy of Mind* (2007) p. 76-77

It is interesting to note that this version of reductive physicalism seems to follow the form of reduction advocated by Carl Hempel and quoted above, who did not want a “strict translation” in which one could, without loss of meaning, translate the concepts in a special science to concepts in a base science, but translation in “a less stringent sense, which does not require that the definiens have the same meaning, or intension, as the definiendum, but only that it have the same extension or application.”⁵⁵ Since it also involves the assumption that the laws in the special science can be derived from the laws of the base science, it has the further reductionist implication that all causality takes place at the level of the base science.

Another form of reductive materialism is *functionalism*. Functionalism comes in different forms and shapes and maybe it is not uncontroversial that functionalism belongs to reductive materialism. At least it is true that functionalists do not claim that the mind is identical to the brain, but rather that the mind is identical to some *function* that the central nervous system carries out. And in this functional description of the mind, it is not the case that they assign mental states *qua* conscious experience a causal role. They leave out or deny the defining characteristic of mental states, their character of being subjective conscious experiences, when they define mentality.⁵⁶ Ned Block talks about “the functionalist program of characterizing mentality in nonmental terms”⁵⁷ and he quotes Sydney Shoemaker who says that “functionalism in the philosophy of mind is the doctrine that mental, or psychological terms are, in principle, eliminable”.⁵⁸ So just as behaviorists, functionalists change the *intension* of mental terms, so as to be able to affirm mental states as having *extension* in a purely material world. Functionalism is thus a form of reduction as translation.

Functionalists approach the mind by first asking “what does it do?” and only secondarily ask “how - and by what structures – does it do it?”.⁵⁹ It’s a bit like asking about the *purpose* of “greeting” – which probably has some more or less universal functions – and only later ask about the specific ways in which people greet in different cultures. So asking what the mind (or a mental state) does is a different question from what it is made of, or how it does it. Therefore functionalists are invulnerable to the argument from multiple realizability (see below): the mind as *function* is not identical to the structure that performs the function. In principle functionalists could be neutral as to whether this function is being executed by something physical or something non-physical, but it seems that all functionalists are materialists in the sense that they would say that the function needs *physical* “realizers”.⁶⁰ Psycho-functionalism is the name of the project of specifying the physical processes that

⁵⁵ Ney, Alyssa “Reductionism” in *Internet Encyclopedia of Philosophy* <http://www.iep.utm.edu/red-ism/#SH1c> retrieved: 2014-08-20

⁵⁶ See: Maslin, Keith T *An Introduction to the Philosophy of Mind* (2007) p. 136 -151; Searle, John R *The Rediscovery of the Mind* (1992) p. 40-45; Block, Ned “Troubles with Functionalism” in *Readings in Philosophy of Psychology* Vol. 1 (1980)

⁵⁷ Block, Ned “Troubles with Functionalism” in *Readings in Philosophy of Psychology* Vol. 1 (1980) p. 296

⁵⁸ Block, Ned “Troubles with Functionalism” in *Readings in Philosophy of Psychology* Vol. 1 (1980) p. 269

⁵⁹ Maslin, Keith T. *An Introduction to the Philosophy of Mind* (2007) p. 122

⁶⁰ Maslin, Keith T. *An Introduction to the Philosophy of Mind* (2007) p. 125

actually perform this function: “the job of spelling out the details of what has come to be called the causal occupant of the functional role”.⁶¹

The typical functionalist description of a mental state such as pain, is to say that it consists of inputs (such as tissue damage) output in forms of pain behavior (such as screaming) and also other mental states (such as a desire to get rid of the pain). But the goal for functionalism is to describe these *mental states* in a non-mental way “it attempts to explicate the nature of mental states in a non-mentalistic vocabulary by reducing them to input/output structures”.⁶² The definition of a mental state, in functionalism, is not that it is a conscious experience, but rather the whole process in which a certain input generates a certain output: “Pain is the whole functional state, characterized in terms of inputs, outputs and relations to other mental states, themselves analyzed functionally.”⁶³ *Analyzed functionally* means analyzed not in terms of their conscious aspects, but in terms of their causal role. This also explains why the typical objections against functionalism try to show that two “systems” can be functionally identical but non-identical in terms of conscious experience, implying that mentality is not reducible to function.

One argument that illustrates this is John Searle’s famous *Chinese room argument*.⁶⁴ The situation is that John sits in a room, with a computer that has the ability to receive input “in Chinese symbols in the form of questions; the output of the system consists in Chinese symbols in answer to the question”.⁶⁵ Someone slips these questions under the door to John and John provides them to the computer who delivers the answers, which John in his turn slips under the door. Now functionally this is indistinguishable from someone sitting in the room actually understanding Chinese, but there is of course a very obvious difference, and the difference is *understanding meaning* not in *performing a function*. Functionalism collapses the distinction between action and behavior, because action is behavior that is intentional, as opposed to a mindless function.

Ned Block also points out that we can think of people with “inverted qualia” - they might have the same qualitative experience when seeing green as others have when they see red. The difference is not in behavior; they would still put all the green things in the same category as someone with a normal qualia-experience, but their conscious experience would be different (not their behavior).⁶⁶ Therefore one cannot exhaustively define mental states in terms of function.

One last version of reductive materialism, not usually understood to belong in this category, is *panpsychism*. In contrast to all other forms of reduction it does not entail an elimination of the mental *qua* mental, but it can still be said to be a form of reductive materialism since it locates mental at the level of the physical. Galen Strawson is, as mentioned earlier, a self-proclaimed physicalist who

⁶¹ Maslin, Keith T. *An Introduction to the Philosophy of Mind* (2007) p 125

⁶² Maslin, Keith T. *An Introduction to the Philosophy of Mind* (2007) p. 136

⁶³ Maslin, Keith T. *An Introduction to the Philosophy of Mind* (2007) p. 137

⁶⁴ Cole, David ”The Chinese Room Argument” in *Stanford Encyclopedia of Philosophy*
<http://plato.stanford.edu/entries/chinese-room/#3> (Retrieved: 2014-08-10)

⁶⁵ Searle, John R. *The Rediscovery of the Mind* (1992) p. 45

⁶⁶ Maslin, Keith T. *An Introduction to the Philosophy of Mind* (2007) pp. 140

argues that physicalists ought to be panpsychists: since a physicalist believes that everything that exists is physical and no conscious being can deny that conscious experiences exist she must admit that consciousness exists as a proper part of the physical domain.⁶⁷ It seems to me that a reduction of the mental to the physical which does not in the end lead to an elimination of the mental, but an “identification” of the mental with something physical, must result in some form of panpsychism, a view of the nature of matter such that it includes mental *qua* mental properties.

In his article *The Myth of Nonreductive Materialism* Jaegwon Kim analyzes the reasons for the “unexpectedly early demise” of reductive materialism in the then most popular form of identity theory.⁶⁸ One was the argument put forward originally by Hilary Putnam known as the “multiple realizability thesis” according to which the same kind of mental phenomena can be realized by more than one kind of brain state. If there was a strict identity between, say, being in pain, and being in a certain brain state, then that would seem to require that one can only be in pain if one is in the brain state which being in pain is supposed to be identical with.⁶⁹ But it seems possible that different brain states could realize the same mental state in different species or different persons. If that is the case then identical mental states are not necessarily identical brain states. Mental states that are alike need not be alike as brain states. Therefore mental states cannot be said to be identical to their associated brain states. This has some connection to the discussion earlier whether a *causal* reduction implies an *ontological* reduction.

Perhaps an analogy can be made between “sentences” and “propositions”. Sentences are linguistic entities which (may) express propositions. Propositions are the semantic content, the meaning of (some) sentences. Therefore the sentence: “I am hungry” is not the same as “Ich habe hunger” although they are identical in their propositional content. Likewise, one could say that two different brain states which both underlie a certain mental state, such as believing it to be Thursday today are related in an analogous way to how sentences are related to propositions.

The other pressure against the identity theory came, according to Kim, from Donald Davidson’s doctrine of the “anomalism of the mental”. Davidson’s basic idea is that neither the analytical behaviorist reduction of the mental to the physical (that talk about mental events is translatable, without loss of meaning, into talk about physical events) nor the so-called “nomological reduction”⁷⁰ that identity-theorists were after, of the mental to the physical, is possible. Davidson’s argument against nomological reduction is based upon the claim that there are no law-like relations between

⁶⁷ Strawson, Galen “Realistic Monism: Why Physicalism entails Panpsychism”

http://www.utoronto.ca/~seager/strawson_on_panpsychism.doc (Retrieved: 2014-08-10)

⁶⁸ Kim, Jaegwon “The Myth of Nonreductive Materialism” in *Proceedings and Addresses of the American Philosophical Association*, Vol. 63, No. 3. p. 32 available at: <http://www.jstor.org/stable/3130081> (retrieved: 2014-08-10)

⁶⁹ Kim, Jaegwon “The Myth of Nonreductive Materialism” (1989) pp. 36-37

⁷⁰ Which they thought of as a form of “ontological reduction” but which Kim shows really is more an expansion of the base theory than a reduction of the target theory. (See below)

mental events, or between mental events and physical events.⁷¹ Davidson argues that any causal relationship between two events “must instantiate lawful regularities”. But mental events are not related in that kind of lawful regularity. Still, mental events do enter into causal relations and causal relations demand nomological (law-like generalizations of cause and effect) relations. Since there are no such nomological relations between mental events on the one hand and between mental and physical events on the other, the mental is “anomalous”:

“the mental is anomalous not only in that there are no laws relating mental events to other mental events but none relating them to physical events either. This meant that no nomological linkages between the mental and the physical were available to enable the reduction of the former to the latter.”⁷²

Davidson’s “solution” to the question of mental causation is to say that mental events are also physical events (each mental event also has physical properties) and it is in virtue of those physical properties that causality takes place. There are at least two problems with this, one that Kim notes and one that he does not seem to note. The first one, which Kim notes is that this seems to make the mental *qua* mental causally inefficacious. Since it is in virtue of the physical properties of the mental event that causality takes place, the mental dimension is inefficacious.⁷³

The second problem, which Kim does not seem to notice, is that if mental events are always physical events and physical events display lawful regularities to other physical events it seems odd why the mental events do not display a lawful regularity in their relation to these other physical events – and the mental events associated with them. That is: why is the mental “anomalous” if a mental event is always also physical and physical events have a lawful relationship to other physical (and mental) events? The only option would be if the mental events are related in a totally unsystematic way to those physical properties. (That would amount to a very strong anti-identity claim, the most heterogeneous relationship thinkable between a particular event’s physical and mental properties: there is no (lawful, but contingent) relationship whatsoever for when a physical description and a mental description are both true of the same event.)

So Davidson’s and Putnam’s arguments worked together to undercut reductive materialism in the form of identity theory. In its place came “non-reductive materialism” and the basic idea is that one can combine ontological physicalism with property dualism.

Non-reductive physicalism argues that the mental is “supervenient” on the brain, it is totally dependent on the brain but it is not ontologically reducible to the brain. The concept of supervenience is not very easy to understand and different writers seem to understand it in different ways. Sometimes a comparison is made by reference to the difference between the properties of the individual dots on a paper and the pattern, or their overall structure. The structure is not a feature of the individual dots, it is a, what Searle would call, “system feature” – it is, however, totally dependent on the dots: it is

⁷¹ Kim, Jaegwon “The Myth of Nonreductive Materialism” (1989) pp. 33

⁷² Kim, Jaegwon “The Myth of Nonreductive Materialism” (1989) p. 33

⁷³ Kim, Jaegwon “The Myth of Nonreductive Materialism” (1989) p. 34

supervenient on - dependent but not reducible to - the dots. The pattern is, however, I would say, explainable and predictable from what one knows about dots (which distinguishes this position from emergence theory).

According to Maslin three aspects define supervenience:

1. The mental (the supervenient level) is not reducible to the physical (the subvenient base).
2. The mental co-varies with and is determined by the physical.
3. Asymmetric dependence: the mental (supervenient level) is dependent upon the physical (subvenient level) but there is no dependence on the mental by the physical or causal arrows from the mental to the physical.⁷⁴

Epiphenomenalism is the view that the mind is a by-product of the activity of the brain but cannot itself exert any causal power. Epiphenomenalism is thus compatible both with non-reductive physicalism and weak emergence. What epiphenomenalism means is that the mental cannot exert causal power, and that the mental is not reducible to the physical. According to Jaegwon Kim, non-reductive physicalism is epiphenomenalism.⁷⁵

According to *emergence theory* mental phenomena emerges out of a material structure (the brain). Mental phenomena are not found among the parts of the brain, they are “system features”.⁷⁶ Emergence theory has garnered quite a lot of interest in the last decades. Emergence theory can be seen as a general ontological framework for understanding how something novel can arise in the process of evolution.⁷⁷ Emergence theory claims that the mind has emerged out of the brain, and is dependent on but irreducible to the brain. In contrast to (at least some versions) of non-reductive physicalism, emergence theory in general denies that the emergence of the mind can be explained by reference to what we know about neurons. The mind is not simply an aggregated effect of many individual parts, it is something new that has come into existence, because of the structure.⁷⁸ *Weak emergence* is sometimes understood as the epistemological claim that we are not able to predict, from our knowledge of the parts, how neurons give rise to the mind, although that in principle should be possible, whereas *strong emergence* argues that it is in principle not possible to give a reductionist account of the emergence of the mind from the brain. Strong emergence also includes the notion that not only new properties can emerge (such as mental life) but also new forms of causality. Strong emergence is therefore compatible with mental causation (qua mental), which seems to imply that the mind can change before the brain changes (which a non-reductive physicalist would deny).

⁷⁴ Maslin, Keith T. *An Introduction to the Philosophy of Mind* (2007) p. 154f

⁷⁵ See especially Kim, Jaegwon “The Myth of Nonreductive Materialism” (1989)

⁷⁶ Beckermann, Ansgar “Introduction – Reductive and Non-reductive Physicalism” in Beckermann, Flohr & Kim (ed.) *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism* (1992) p. 16

⁷⁷ Clayton, Philip *Mind & Emergence: from Quantum to Consciousness* (2004) s. 39

⁷⁸ Beckermann, Ansgar “Introduction – Reductive and Non-reductive Physicalism” in Beckermann, Flohr & Kim (ed.) *Emergence or Reduction? Essays on the Prospects of Nonreductive Materialism* (1992) p.

Cartesian substance dualism argues that the mind is an independently existing substance, which somehow is connected to the brain and maybe also needs the brain in order to have any conscious episodes.⁷⁹ The mind can be affected by the body and the brain, and can affect it in turn. An interesting question is if it is the soul that is responsible for our consciousness or the brain, if the soul needs a (functioning) brain in order to have any conscious experiences. It seems as if, of the above mentioned alternatives, only strong emergence, Cartesian substance dualism and the pan/localpsychism version of reductive materialism allows for mental causation qua mental.

In philosophy of mind some central issues have framed the discussion in the last decades, and they have been presented in some seminal papers. Among them are Thomas Nagels *What Is it Like to Be a Bat?* in which he argues that a fundamental aspect of conscious experience is its subjective and personal character: a conscious organism is an organism that experience the world in a certain way.⁸⁰ Another important essay is David Chalmers paper *Facing Up to the Problem of Consciousness* in which he distinguishes the “easy problems” of consciousness from the “hard problem” of consciousness.⁸¹ The easy problems (relatively, that is) consist in, for instance, trying to map out which areas of the brain that is responsible for different functions. The hard problem is to try to describe how something physical can give rise to the mental. Jaegwon Kim’s *The Myth of Nonreductive Materialism* – referred above – has also been influential for the debate.

Chapter 2. Physicalism, mentality and mental causation

2.1 Introduction

Most people would agree that there are material, or physical, objects in the world. Some argue that there only exists physical objects, and they are called physicalists. A distinction is usually made between *strict* or *reductive physicalism* according to which not only are there only physical objects: all *properties* that exists are also physical in nature, and *non-reductive* or *weak* physicalism according to which non-physical properties might exist, which “supervene” on – that is, they are distinct from but dependent on and determined by - physical objects and processes.⁸² A further distinction can be made between non-reductive physicalism and emergence theory: according to emergence theory these physically irreducible properties can exhibit (new forms of) causal powers.⁸³ The dominant position –

⁷⁹ This seems to be the position of Richard Swinburne. See his “Själen behöver en hjärna för att fungera” in Hasker, Peterson, Reichenbach, Basinger (ed.) *Religionsfilosofiska texter* (1999) p. 468

⁸⁰ Nagel, Thomas “What is it like to be a bat?” *The Philosophical Review*, Vol. 83, No. 4 Oct., 1974 Available at: <http://www.jstor.org/stable/2183914> (retrieved: 2014-08-10)

⁸¹ Chalmers, David “Facing up to the Problem of Consciousness.” In *Journal of Consciousness Studies* 2: 200-19 (1995) <http://consc.net/papers/facing.html> (retrieved: 2014-08-30)

⁸² Crane, Tim; Brewer Bill “The Mental Causation Debate” in *Proceedings of the Aristotelian Society Supplementary* Volume 69, 1995 available at: <http://sas-space.sas.ac.uk/218/1/The%20mental%20causation%20debate.pdf> (retrieved: 2014-08-10)

⁸³ Clayton, Philip *Mind and Emergence: From Quantum to Consciousness* (2004) p. 49-54 Probably supervenience, understood as total determination of the mental by the physical is an inadequate description for

the orthodox position as it is often called – within philosophy of mind is non-reductive physicalism, according to which mental properties are real and irreducible to physical properties, although they are totally dependent on and determined by their physical base.⁸⁴

Jaegwon Kim has for the last 25 years been arguing against this orthodoxy in favor of reductive physicalism, and his main argument is that non-reductive physicalism is not compatible with mental causation.⁸⁵ Jaegwon Kim argues for reductive physicalism, and he thinks that this position saves mental causation. It is my intention in this chapter to present Kim's criticism against non-reductive physicalism, which I believe is sound, and analyze his solution, which I believe in fact does not give room for mental causation, as Kim (and I) understand mental causation. But before presenting Kim's argument I want to give a brief description of what we mean by physicalism, mentality and mental causation.

2.2 What do people (in general and Jaegwon Kim in particular) mean by physicalism?

Most people have some understanding of what “physical” means and would probably exemplify it by reference to such properties as having mass, being extended and having electrical charge. These are paradigmatic and non-problematic examples of physical properties. We can also think of phenomena – real or not – which do not seem to fit in this category of physical objects or properties. Examples would be prime numbers, angels, God(s), propositions, and mental phenomena such as thoughts, feelings, desires and beliefs. Although the ontological status of mental phenomena is debated - physical or not? - everyone would agree that they are not as easily described as physical as for instance atoms are. They are certainly not the paradigmatic cases of physical objects, they are, as Jaegwon Kim says, *prima facie* non-physical.⁸⁶ So there are objects that everyone who thinks there are physical objects at all would agree are physical (such as stones) there are other phenomena that, if they exist, clearly are non-physical (such as laws of logic) and there are other objects which seem to have both physical and non-physical properties, such as conscious human beings.

But against what kind of “yardstick” do we judge phenomena to see if they are physical or not, or perhaps find that they are hard to decide whether they are physical or not? According to Daniel Stoljar there are basically two strategies.⁸⁷ One is by reference to the conceptual resources of physical theory and one is by reference to the properties that the objects which are paradigmatic examples of “physical objects” have. If defined by reference to physical theory, everything is physical that is a part of the conceptual resources of (current or a future ideal and therefore complete) physics and if defined

understanding the relationship between mental and physical properties in emergentism, since it seems to imply that all causation takes place at the subvenient level. (See Jaegwon Kim (2005) p. 36-39)

⁸⁴ Crane, Tim “The Mental Causation Debate” (1995) p. 1 (Crane calls non-reductive physicalism for “composition physicalism”, but is explicit that what he means is the same as what is traditionally called non-reductive physicalism.)

⁸⁵ To my knowledge, the first formulation of this argument was his presidential address before the American Philosophical Association, titled “The Myth of Nonreductive materialism” (1989)

⁸⁶ Kim, Jaegwon *Mind in a Physical world* (1998) p. 96

⁸⁷ Stoljar, Daniel “Physicalism” in *Stanford Encyclopedia of Philosophy* (<http://plato.stanford.edu/entries/physicalism/> (retrieved: 2014-08-25))

by reference to physical objects, a property is physical if it is a property that would be included in a full description of the intrinsic nature of paradigmatic physical objects. (It is hard to see, however, how something that is needed for a full description of the intrinsic nature of paradigmatic physical objects, could fail to be part of physical theory.)

However, as Jaegwon Kim says, there is no consensus about how to formulate physicalism⁸⁸ and there are problems with both of these strategies. For instance, it is often claimed that if physicalism is defined by reference to a future physics, “physicalism” can become an empty concept: who knows what the future has in store for physical theory? Physicalism is a modern form of materialism, but our current understanding of the nature matter is quite different from the materialism of the 17th century when matter was considered to be “solid, inert, impenetrable and conserved, and to interact deterministically and only on contact”.⁸⁹ According to Crane and Mellor this turned out to be wrong in every aspect (at least if this is thought to be necessary features of matter rather than true only given certain conditions) one of the most radical changes being that matter seemingly *can* interact at a distance.⁹⁰ And Popper, making the same argument that “materialism transcended itself” in the very process of studying matter, points out that “matter turns out to be highly packed energy, transformable into other forms of energy; and therefore something of the nature of a *process*”.⁹¹ Daniel Stoljar even argues that panpsychism is compatible with physicalism:

“After all, the fact that there are *some* conscious beings is not contrary to physicalism — why then should the possibility that *everything* is a conscious being be contrary to physicalism?”⁹²

One self-proclaimed physicalist, Galen Strawson, quite forcibly argues that physicalists ought to be panpsychists. In his essay:

Monism: Why Physicalism entails Panpsychism he argues that since conscious experience is an undeniable aspect of reality, and since physicalists think that everything that exists is physical, physicalists must admit that the physical also has mental aspects, a position that is generally labeled panpsychism.⁹³ I take it that the defining character of the position known as “panpsychism” is that at least some matter has, among its intrinsic properties, mental properties, and not that *all* matter has mental properties. The crucial idea is that mental aspects do not supervene on or emerge from material objects, it is as much a part of the properties of individual material objects as mass or electrical charge.

But clearly there is a tension between physicalism and panpsychism. We can of course just define “physical” as anything that happens to be real, and then whatever exists is physical, including

⁸⁸ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 33

⁸⁹ Crane, Tim; Mellor, D. H. “There is no question of physicalism” *Mind*, New Series, Vol. 99, No. 394 (Apr., 1990) p. 186 available at: <http://www.jstor.org/stable/2254959> (Retrieved: 2014-08-30)

⁹⁰ Gravity and quantum entanglement are usually mentioned as examples of interaction at distance.

⁹¹ Popper, Karl & Eccles, John *The Self and Its Brain* (1977) p. 7

⁹² Stoljar, Daniel “Physicalism” in *Stanford Encyclopedia of Philosophy* <http://plato.stanford.edu/entries/physicalism/> (Retrieved: 2014-08-25)

⁹³ Strawson, “Galen Realistic Monism: Why Physicalism entails Panpsychism” *Journal of Consciousness Studies*, Vol. 13, No. 10-11, October/November 2006 http://www.utsc.utoronto.ca/~seager/strawson_on_panpsychism.doc

mental properties. But then it does not imply anything about your ontology to say that you are a physicalist, and if physicalism implies no limitation to what kind of phenomena you can affirm as real, your physicalism lacks definition and therefore meaning. In so far as it means something to say that one is a physicalist, it also implies a limitation to your ontology.

And most physicalists would not be comfortable with panpsychism, and the reason of course is that they implicitly or explicitly think that our physical theory and/or our understanding of the nature of “paradigmatic physical objects” such as atoms is at least somewhat complete, it actually gives you a limitation to what kind of phenomena and properties you can affirm as real and part of the furniture of the universe. Our understanding of the intrinsic nature of matter will not be so radically expanded as to include what we today would call mental properties. That is why the typical physicalist strategy towards mental phenomena is either to try to describe them in a non-mental way or accept them as irreducible to physical properties but argue that they supervene on and are determined by something physical.

In this essay it is Jaegwon Kim’s understanding of physicalism that is of most interest. As far as I can see Kim does not define physicalism – he explicitly says that there is no agreed definition - but he provides some clues of what he means by the term.⁹⁴ He describes physicalism as “the thesis that bits of matter and their aggregates exhausts the content of the world”⁹⁵ and that these bits and aggregates behave “in accordance with physical law”.⁹⁶ If the expressions “bits of matter” and “physical law” are to have some meaning, it must be with reference to our current understanding or an understanding that is not too different from our current understanding.⁹⁷

And when one sees how Jaegwon Kim goes about when trying to make sense of mental phenomena within a physicalist worldview, it is clear that he does not appeal to a radically expanded physics. So he tries to understand mental phenomena as things that are explicable given our current physical theory (although he admits that not all mental phenomena are possible to accommodate within physicalism, and that physicalism therefore does not survive in its entirety).⁹⁸ So to summarize: Kim’s physicalism is a form of physicalism according to which whatever is real must be possible to account for given the conceptual resources that current physical theory gives us – or at least not imply a too radical expansion or transformation of physical theory.

One concise and helpful definition of physicalism or naturalism – these term are often used interchangeably –is provided by J.P Moreland in his article *The Argument from Consciousness* in *Blackwell Companion to Natural Theology*. Moreland points out three central tenets of naturalism:

⁹⁴ When trying to define non-reductive physicalism Kim acknowledges that neither reduction nor physicalism has a well-defined meaning, making non-reductive physicalism hard to define. *Physicalism, or Something Near Enough* (2005) p. 33

⁹⁵ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 3

⁹⁶ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 7

⁹⁷ “Not too different” is of course a loose criteria, reflecting the vagueness of physicalism as a theory of reality.

⁹⁸ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 31

- I. An epistemic attitude, either understood as the claim that science is the only road to knowledge, or that science is at least extremely privileged as a way of getting knowledge. What this implies, among other things, is that the part of reality that is exhaustively describable from a third-person perspective tends to be seen as more real than things such as consciousness which are not.
- II. The “naturalist grand story”. The naturalist grand story is the natural scientific theories of how everything got into being, from the Big Bang and onwards. John Searle, whom Moreland is indebted to in this account, writes, related to this, that the “modern world view” is constrained mainly by the atomic theory of matter and evolutionary biology.⁹⁹
- III. The naturalist ontology. There is a distinction here between weak and strong physicalism, according to the latter everything that exists must itself be physical and according to weak physicalism there might be properties that are not themselves physical but which supervene on something physical. The basic task for the physicalist is to try to give an account of how a phenomenon arose given the constraints that the “naturalist grand story” provides - among them: the atomic theory of matter and evolutionary biology.

The distinction between “serious metaphysics” (as the materialist Frank Jackson calls it) and “shopping-list metaphysics” (as Moreland calls it) is relevant in this context. The point of “serious metaphysics” is to try to locate phenomena within a naturalist ontology, that is, to explain how they can be and how they got into being, given the constraints of the “naturalist grand story”. According to Frank Jackson, a phenomenon must either be possible to locate in “the grand story” or else eliminated from one’s ontology. Although there is something that is valid in this strategy there is also something problematic with it. If the lack of an adequate *explanans* can dictate whether an *explanandum* is real or not, you might be arguing in a circle: “x does not exist” – “how do you know?” – “I cannot account for it given the constraints of my worldview”. At some point it might be necessary to revise one’s ontological commitments. The point of a worldview is to make sense of one’s experience and observations of reality, not to dictate in beforehand what is real or not. If this strategy is pursued consistently, one’s worldview would be incapable of change as one makes new and unexpected observations.¹⁰⁰

Although Moreland argues that naturalists ought to be “strict physicalists” it should be mentioned that many take it that a tension exists between strict physicalism, and the integrity of the so-called special sciences, such as evolutionary biology, which is *also* thought to be part of the grand story. The reason is that if everything is reducible to microphysical causation there is actually no real causal work to be done at the level of cells or organisms, and therefore no explanatory work to be done

⁹⁹ Searle, John *The Rediscovery of Mind* (1994) p. 86

¹⁰⁰ An interesting example could be the study of altruism. According to Daniel C. Batson who has studied this a lot, he did not believe that genuine altruism existed when he began his studies. He has since then changed his mind, due to his empirical findings, but finds the scientific community in general to have the attitude that it is, in some *a priori* way more “scientific” to say it does not exist. I believe this is connected to the (perceived) difficulties of locating altruism within the naturalist grand story.

in the disciplines that study them, in addition to the explanatory work done by the disciplines that study the (only) causality that takes place (at all) at the microphysical level. As Moreland writes:

“As typically presented, the hierarchy entails the causal closure of the basic microphysical level along with the ontological dependence of entities and their activities at the supervenient levels on entities and their activities at that basic level.”¹⁰¹

This has the further consequence that the *hierarchy* is only phenomenological. It is only a matter of perspective: ontologically speaking there is only one level, the microphysical. There is a huge tension between that view and the view that there are non-redundant explanations in different fields of science and irreducible objects of study within for instance biology (not to speak of psychology).¹⁰²

However, evolutionary biology is often taken to be a non-problematic part of the naturalist grand story, whereas, given strict physicalism it is not.¹⁰³ In any case, a constraint for properties of living organisms given the naturalist grand story is that they should be possible to explain given evolutionary biology which implies that they:

“exist and are what they are because they contributed to (or at least did not hinder) the struggle for reproductive advantage, more specifically, because they contributed to the tasks of feeding, fighting, fleeing and reproducing”.¹⁰⁴

This is important for my essay, because it is my intention to show that if mental phenomena are to be located within the naturalist grand story, they must contribute to these tasks, *qua* mental, which they cannot if they are epiphenomenal.

2.3 Mental phenomena

What do we mean by mental phenomena? Peter Gärdenfors, professor of Cognitive Science at Lund University, writes in his book *Hur Homo blev Sapiens: Om tänkandets evolution* about the different aspects of our mentality, some of which we share with other animals, and some of which are specifically human.

First I would like to make a general remark that when I speak of mental phenomena, I am interested - if they are not the same thing - in *conscious* phenomena. And *consciousness* as such is impossible, I think, to define purely from a third-person perspective so as to give someone who does not have a first-hand experience of consciousness full knowledge of what it means to be conscious (for the moment ignoring the fact that in order to be able to understand anything at all, you would have to be conscious). The point is that you cannot define consciousness, you can only make a conscious being more aware of and attentive to different aspects of her conscious life, and that is the purpose of

¹⁰¹ Moreland, James Porter “The Argument from Consciousness” in *The Blackwell Companion to Natural Theology* (2012) p. 287

¹⁰² And this tension is the subject-matter of the debate between Ned Block and Jaegwon Kim on whether causal powers “drain away” if the supervenience argument is successful. See Kim (2005) p. 57 - 69

¹⁰³ And arguably also weak physicalism since supervenience-physicalism does not allow for the emergence of causal powers.

¹⁰⁴ Moreland, James Porter “The Argument from Consciousness” in *The Blackwell companion to Natural Theology* (2012) p. 285

this chapter. The goal is to get a clearer understanding of what we mean by mentality and what it is that is supposed to be causally active in mental causation.

In the first category Gärdenfors lists bodily sensations, a phenomenon we share of course with a lot of other species. These are the conscious experiences we get through our bodily senses, that is, through our hearing, smell, vision, taste and sensation.¹⁰⁵ These are the five classical senses; today these are usually complemented by nociception - our ability to feel pain, and proprioception - our awareness of the location of our own body and our body parts. A distinction that Alvin Plantinga makes between “indication” and “conscious belief” is helpful here. All living organisms get some information from their environment, and this is also true of plants. They get “indications” about their environment and their behavior will change, automatically and without any conscious belief or experience on their part.¹⁰⁶ The organism’s behavior can, if it is governed by indication rather than belief, be described as a function that changes after it receives information about the environment. This is true also of some human behavior, such as when our hands automatically draw away from a hot stove and we first later experience the pain that the heat causes – first there was indication and later there was conscious experience. Here the aforementioned distinction in sociology between behavior and action is applicable: actions are intentional, voluntary and deliberate, whereas behavior is more reflexive and automatic.¹⁰⁷

The process in which the physical stimulus is transformed to nerve impulses (from, for instance, mechanical energy in the case of touch) is called “transduction”.¹⁰⁸ It could be said that explaining the “hard problem of consciousness” (that is explaining how something material can give rise to conscious experiences) would consist in giving an account of a second form of transduction: how are electrical nerve impulses transformed to mental experience?¹⁰⁹

In the next category Gärdenfors places “perception” which is a form of interpreted sensations. Here we have the first rudimentary forms of thinking. This is connected to what is called, in cognitive psychology, “top-down processing” which takes place whenever a stimulus is interpreted by the conscious organism “in light of existing knowledge, concepts, ideas and expectations”.¹¹⁰ So there is a certain cognitive aspect to perceptions that may not be there regarding sensations. Perceptions involve some sort of theory about what causes the sensations, some rudimentary understanding of the cause of sensation. Therefore Gärdenfors describes sensations as answering the question: what is happening to my body, whereas perceptions try to answer: what is happening in the world? Trying to understand the cause(s) of sensation is the first role of perceptions.¹¹¹

¹⁰⁵ Holt, Nigel *Psychology: The Science of Mind and Behavior* (2012) p. 142 - 158

¹⁰⁶ Plantinga, Alvin *Where the Conflict Really Lies: Science, Religion and Naturalism* (2011) p. 328

¹⁰⁷ Engdahl, Oskar; Larsson, Bengt *Sociologiska perspektiv* (2006) pp. 31

¹⁰⁸ Holt, Nigel *Psychology: The Science of Mind and Behavior* (2012) p. 142

¹⁰⁹ Conveniently enough this problem is often not even mentioned in most introductory textbooks of psychology, giving the unreflective student the impression that there is nothing that needs to be explained here.

¹¹⁰ Holt, Nigel *Psychology: The Science of Mind and Behavior* (2012) p. 163

¹¹¹ Gärdenfors, Peter *Hur Homo blev Sapiens: Om tänkandets evolution* (2000) p. 36

The next category is attention. Our attention determines what part of the external (or internal) world that is most psychologically present for the conscious subject at the moment. There is a lot of information that is being received and processed by the brain that is not attended to: Gärdenfors likens the brain to a construction site in which a lot of work is being done, and *attention* can be seen as that part of the site which is being lit and where some especially important or hard work is being done.¹¹²

In the next category Gärdenfors puts emotions, such as fear, happiness, anger and the like. They often function as guides to action, especially when a fast response is needed. That's also the reason why we sometimes say that we "overreacted": we emotionally interpreted the situation as more alarming than it, upon further reflection, turned out to be.

Memory is a very important aspect of thinking and a prerequisite for learning. There is a distinction between implicit and explicit memory, where explicit memory is that kind of memory that we can articulate. Some patients whose hippocampus has been damaged have a hard time learning new things, since the hippocampus is involved in the formation of new memories, although they sometimes seem to be able to acquire new implicit memories. There is another distinction between procedural, semantic and episodic memory. Of these procedural can be said to be implicit. It is the kind of memory that consists in learning how to, for instance, ride a bike or tie one's shoes. Semantic memory is about learning facts and episodic memory is the memory of events. As far as we know only humans have episodic memory, which makes it possible for us to "travel in time" and re-experience events, and also travel forward in time and see our own mortality.

One interesting distinction that Gärdenfors quotes is the difference between different kinds of organisms, depending on to what extent they can learn and integrate new information. This is the distinction between *Darwinian*, *Skinnerian*, *Popperian* and *Gregorian* creatures, owing their name to the different scientists who are famous for studying the respective capacity. Darwinian creatures are those creatures whose only information storage is in the genetic material, in the DNA. Plant's are "programmed" to seek sunlight and try to get water from the soil. The only way in which "learning" can occur is through mutations and natural selection: learning takes place, so to speak, over generations. It might be the case that the offspring, due to mutations, get characteristics that the "parents" did not have, and that if they turn out to be helpful in the environment, it will spread in the population.

A Skinnerian creature is a creature that does not only have a genetically programmed repertoire of behavior, it can also learn to associate things in the environment as good and as bad. It will only happen as a result of experience though. It is called "conditioning" and these creatures are called Skinnerian, of course, because B F Skinner studied this process. One important difference between the Skinnerian and the Popperian creature is that the Skinnerian only has perceptions – that is interpretations – of things that it has empirical, that is sense perceptive, contact with, in its immediate surroundings – a Skinnerian creature therefore only lives here and now.

¹¹² Gärdenfors, Peter *Hur Homo blev Sapiens: Om tänkandets evolution* (2000) p. 17

A creature that is called Popperian has an inner representation, a cognitive map, of the world in which hypothetical scenarios can be tried out. This demands representations that are not merely interpretations of one's sensations, but also presupposes a mental life that to some extent exists independently of stimulation. So "object constancy" – the realization that things are real even when they are not present – requires mental representations of objects, a representation that can be held in the mind independently of sensation. A Popperian creature is a creature who does not need to get the experience that something is bad in order to learn that is potentially bad, but instead can simulate the action in its internal world, in its internal representation of the external world. This is a great advantage because it saves the organism a lot of trouble, and it can become proactive and not only reactive. It can of course also be applied regarding potentially good things and actions. Although a Popperian creature is not necessarily one who has the cognitive abilities connected to what Jean Piaget called the "formal operational stage" (thought to begin when humans are about 12 years old) one could say that an organism that has entered the "formal operational stage" is a Popperian creature who has maximized the benefits of having inner representations, since she is able to manipulate them independently of experience:

The formal operational thinker has the ability to consider many different solutions to a problem before acting. This greatly increases efficiency, because the individual can avoid potentially unsuccessful attempts at solving a problem. The formal operational person considers past experiences, present demands, and future consequences in attempting to maximize the success of his or her adaptation to the world.¹¹³

Lastly Gärdenfors talks about Gregorian creatures that not only have an inner world, but also have the ability to receive knowledge from other's inner world, through language and culture.

Self-consciousness is not only an awareness of one self as a body, but an awareness of one self as a subject, with an inner world and a certain first-person perspective. Gärdenfors argues that the experiments in which Chimpanzees have shown that they understand that the body that they see in a mirror is their own body is not evidence of self-consciousness, but of body-awareness, awareness of oneself as a physical object rather than as a conscious subject.¹¹⁴ It is also interesting to note that Gärdenfors argues that consciousness of other having an inner world precedes the self-reflective awareness of one's own inner world. Being aware of other having a perspective opens up possibilities of cooperation (and community) which is an evolutionary advantage, but it also implies the ability to realize that you can manipulate others and that they can manipulate you. Both of these phenomena probably exerted selective pressures towards greater abilities of social cognition. One possible biological advantage of self-consciousness (which Gärdenfors does not mention) is that it makes you aware of the fact that you might be wrong, that you have a limited perspective.

So, all organisms are to some extent sensitive to information from the environment, otherwise the organism would not be able to respond in a way that was beneficial for the organism. But in some

¹¹³ Cherry, Kendra "Formal Operational Stage of Cognitive Development"
<http://psychology.about.com/od/piagetstheory/p/formaloperation.htm> (retrieved: 2014-08-25)

¹¹⁴ Gärdenfors, Peter *Hur Homo blev Sapiens: Om tänkandets evolution* (2000) p. 115

animals this information also becomes conscious. The issue in mental causation is whether this conscious awareness is merely an epiphenomenon so that there is a cause both for our awareness and our action, or whether conscious mental acts can be causes of behavior, perhaps as an intermediary variable.

2.4 Mental causation

In his presidential address *The Myth of Nonreductive Materialism* Jaegwon Kim criticizes Donald Davidson's view on mental causation. For Davidson all mental states also have physical properties – it is more clarifying to put it the other way around: mental properties supervene on brain states – and it is in virtue of those physical properties that causality takes place.¹¹⁵ Kim comments:

It is true that in a Davidsonian domain, all events are physical; that is every event has some physical property. But when I say that mental events cause physical events something stronger is intended, namely that an event, *in virtue of its mental property*, causes another event to have a certain physical property.¹¹⁶

Kim acknowledges that this is a stronger view of causality than simple Humean counterfactual dependence, and argues (rightly in my view) that our “core idea of causation is more intimately tied to generative/productive causation than to counterfactual dependence”.¹¹⁷ To put it another way: the reason that counterfactual dependence between A and B holds, is because there is a causal relationship of the productive/generative kind between A and B.¹¹⁸ And as Kim says in a response to Barry Loewer who argues that “productive causation” is not a scientifically sound concept, we need this notion of causation in order to save human agency:

My first thought is that if the situation *is* indeed as Loewer describes it and there is no productive causation anywhere, then there is no mental causation anywhere and the fear, common among philosophers and nonphilosophers alike, that physicalism, or physics, takes away human agency, would be amply warranted. Why do we care about mental causation? Because, first and foremost, we care about human agency. To save agency, however, we *need* the productive concept of causation; we want agents, in virtue of the beliefs and desires and intentions they hold, to cause their limbs to move in appropriate ways and thereby produce changes in their physical surroundings.¹¹⁹

It seems to me that thinking of causality merely in terms of counterfactual dependence dissolves the distinction between correlation and causality. The phenomenon of daylight always follows after the dark of the night, but the latter is not the cause of the former, rather there is a common cause (a “hidden variable”) to both phenomena. So a thick notion of causality is needed, and causation as production is the one I will adopt when discussing whether mental causation is real. And the causation

¹¹⁵ Kim, Jaegwon *The Myth of Nonreductive Materialism* (1989) p. 43

¹¹⁶ Kim, Jaegwon *The Myth of Nonreductive Materialism* (1989) p. 43 Italics in original

¹¹⁷ Kim, Jaegwon *Physicalism, Or Something Near Enough* (2005) p. 18 note 12

¹¹⁸ It is by many argued that counterfactual dependence should be seen as one of several epistemological criteria of causality, not the ontological meaning of causality itself. See “Orsak/verkan” in *Filosoflexikonet* (1988) p. 4

¹¹⁹ Kim, Jaegwon “Responses” *Philosophy and Phenomenological Research* Vol. 65, No. 3, November 2002, p. 675 available at: <http://onlinelibrary.wiley.com/doi/10.1111/j.1933-1592.2002.tb00231.x/pdf> (Retrieved: 2014-08-10)

will be a form of causation in which one or several conscious phenomena, say a belief and a desire, are the productive causes. That is what makes it mental causation.

2.5 The basic argument against non-reductive physicalism

According to Kim, non-reductive physicalism can be understood as the following three theses:

- I. **The mental *supervenes* on the physical.** Although “supervenience” does not have a clear definition, it is generally understood as implying a determinative relationship between the supervenient phenomenon by its subvenient base. An example can be made by reference to the structure that a number of dots on a paper forms (supervenient level) and the individual dots (subvenient base). The structure is not the same as the dots, but it is totally determined by the dots. It is dependent on and determined by the dots. It is a one-sided dependence, asymmetric, so that the base determines the top, but not the other way around, and there is no interaction between them.
- II. **The non-reducibility of the mental to the physical.** By reference to the dots: the structure is not reducible to the dots, although it is dependent on them. For me it seems as if non-reducibility is a necessary part of supervenience (because a thing does not supervene on itself), but Kim understands this as a separate thesis from the supervenience thesis. This is a bit strange since according to Kim to reduce something to something else is actually to identify it with that something¹²⁰ and it does not make sense to talk about something supervening on itself.
- III. **The mental is causally efficacious.**¹²¹ As explained earlier, the notion of causality we are interested in is a stronger understanding of mental causation than the Davidsonian understanding according to which a mental event will also have physical properties and those physical properties are the only that are causally efficacious. It is the productive/generative understanding of causality.

The basic idea in Kim’s criticism is that it not possible to hold these three theses if one is a physicalist. The reason is that as all physicalists must accept the causal closure of the physical domain¹²² and the mental is not reducible to the physical, according to non-reductive physicalism, it follows that mental causation would be a form of mental-to-physical causation. Or to put it another way: if the mental is not reducible to the physical and it can be causal efficacious, it would follow that some physical events (such as for instance my writing this sentence) would have a non-physical cause.

And that violates the principle of the causal closure of the physical domain, according to which “if a physical event has a cause that occurs at t , it has a physical cause that occurs at t ”.¹²³ Two things are noteworthy with this specific formulation of the causal closure principle: first that Kim allows

¹²⁰ Kim, Jaegwon, *Physicalism or Something Near Enough* (2005) p. 34

¹²¹ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 35

¹²² Kim, Jaegwon *Mind in a Physical World* (1998) p. 40

¹²³ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 43

there to be physical events without causes, it is just that if there are causes for an event, they must be physical.¹²⁴ Secondly that it would not be enough, for Kim, to just say that an event must have a physical cause, it must have a physical cause “at *t*” because otherwise it might give room for mental causes which were part of a chain of both physical and mental causes leading to the physical event in question:

“A simpler statement of causal closure in the form “If a physical event has a cause, it has a physical cause” will not do; given the transitivity of causation, the requirement would be met by a causal chain consisting of a physical effect caused by a mental cause which in turn is caused by a physical cause.”¹²⁵

And the supervenience relationship between the mental and the physical implies that in order to have mental-to-mental causation you must change the physical subvenient base in order to change the mental supervenient level. Given non-reductive materialism, you cannot have mental-to-mental causation unless you have mental-to-physical causation. Kim comes to the conclusion that in order to save mental causation we should not give up causal closure, instead we must reduce mental properties to physical properties: “if mental properties are to be causally efficacious, they must be physically reducible”.¹²⁶

To summarize: Kim’s critique of non-reductive physicalism is that if you accept the causal closure of the physical domain, as you must do if you are a non-reductive *physicalist*, and the irreducible character of mental properties, as you must do if you are a *non-reductive physicalist* you cannot have mental causation.

It is now time to look at Kim’s solution to this problem. As can be easily seen, there are two ways out of this dilemma: either you give up the causal closure of the physical domain, or you try to reduce mental properties to physical properties. Kim, being a good and faithful physicalist, takes the second route. However, since the motivation for Kim to do this is to save mental causation *qua* mental, that is: “that an event, *in virtue of its mental property*, causes another event to have a certain physical property”¹²⁷ the task is not only to reduce mental properties to physical, it is to reduce them without eliminating their causal efficacy *qua* (what is traditionally called) mental.

2.6 Kim’s project of reducing mentality

In this section I will first summarize how Kim criticizes Nagelian reduction for being an expansion of the base theory rather than a reduction of the target theory. Secondly I will present Kim’s view of functional reduction as the right form of reducing mentality and I will show that Kim’s functional reduction is not consistent with his motivation for reducing the mental to the physical in the first place, namely to safeguard mental causation *qua* mental.

Kim discusses different forms of reduction in chapter four in *Physicalism, or Something Near Enough* and he begins with a discussion on the relationship between reduction and reductive

¹²⁴ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 16

¹²⁵ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 43 note 10

¹²⁶ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 5

¹²⁷ Kim, Jaegwon “The Myth of Nonreductive Materialism” (1989) p. 43 Italics in original

explanation, which parallels my discussion above on the relationship between ontological and causal reduction. Kim notes that many non-reductionists argue that reductive explanations (causal reductions) might still be available even if you will not be able to carry through an ontological reduction, that is, an *identification* of the phenomenon with its underlying causes.¹²⁸

Kim's basic critique of the Nagelian form of reduction is that it is neither an ontological reduction nor a reductive explanation: it is merely a description of the correlation that stands in need of explanation (and, for the reductive physicalist: in need of ontological reduction).¹²⁹ The Nagelian form of reduction is one in which you have phenomena in the target theory that are correlated with phenomena in the base theory. Let us say that "pain" is a phenomenon in the target theory (the theory to be reduced) and that it is correlated with a phenomenon in the base theory, for instance stimulation of certain nerve fibers, let us call them (as is common in the literature) c-fibers. Since we cannot deduce "pain" from our neurophysiological theory, the Nagelian reduction is accomplished by postulating "bridge laws" that purports to transform the correlation into a causal relationship: pain occurs if and only if c-fibers are stimulated, in virtue of bridge laws. The problem is that this does not reduce pain to something neurophysiological as much as it introduces pain in the base theory. "Pain" is not a property that originally occurred in the base theory, but you now have a law in the base theory in which "pain" occurs as a "primitive" (that is, not reduced) property:

"This means that that the pain-Cfs correlation law is now one of the axioms of the expanded base theory, and that this new base theory no longer is a pure neurophysiological theory – its primitive predicates include "pain", and among its axioms is a law about pain."¹³⁰

In order to get a reduction, you must "derive the laws being reduced *solely from the explanatory resources available in the base domain*"¹³¹ and this is not done in a Nagelian reduction. It is an expansion of the base, and for that reason not a reduction of the target theory, but an expansion of the base theory:

"In a Nagel reduction, however, the bridge laws connecting the two domains enter the picture as indispensable supplementary assumptions, and this means that the base theory has been expanded both in its ideology (via addition of new predicates/concepts) and in its ontology (via new properties). Moreover, these a posteriori bridge laws are now among the basic laws of the expanded base theory."¹³²

To say that c-fibers cause pain in virtue of something called "bridge-laws" is not the same as identifying c-fibers with pain and the identity theorists were clear that they wanted to get away from expressions like "causing" or "correlating with" – they wanted an identity. Kim concludes therefore that Nagelian reduction is not a reduction of the phenomenon at all, and neither is it a causal *explanation* of the phenomenon: it is just a statement of the observed correlation, together with an assumption that there is some causal mechanism that could explain the correlation:

¹²⁸ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 95

¹²⁹ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 97-100

¹³⁰ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 100

¹³¹ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 100 emphasis in original

¹³² Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 100

“In using the bridge laws as auxiliary premises of reductive derivations, the Nagelian reductionist is simply assuming exactly what needs to be derived and explained”¹³³

In general terms, Kim formulates the following constraints on what can be a candidate for a reductive explanation:

“The explanatory premises of a reductive explanation of a phenomenon involving property F (e.g. an explanation of why F is instantiated on this occasion) must not refer to F.”¹³⁴

The form of reduction that Jaegwon Kim advocates is functional reduction and it involves in Kim’s version three steps.¹³⁵ The first step is to define in more technical terms the phenomenon you want to reduce, and it is called “functional reduction” because the definition should be in terms of the causal work - “the function” - the property is supposed to perform. If we exemplify it with for instance “reproduction” the definition could be “a cellular process leading to the production of a complex physical entity that is similar to the original entity”.

The second step is to find the specific realizers of this function. If the function to be reduced is transmission of genetic material, among the realizers will be the DNA-molecule. The third step is to formulate a theory, in the language of the base-theory, how this function, defined in step one, is carried out by the physical realizers defined in step two. What survives in this reduction is the process of reproduction and the physical realizers of that function. Nothing essential seems to be lost from our understanding of reproduction in this reduction. As David Chalmers says, if someone would complain that we have now given an account of how “a cellular process leads to the production of a complex physical entity that is similar to the original entity” but not explained *reproduction* we would be impatient, because that is what reproduction means.¹³⁶

The crucial thing is that the technical definition of “reproduction” did not leave out any essential meaning in the word “reproduction”. Applied to the mind, the functional reduction involves three steps:

- I. *Functionalize* the mental state, by identifying the causal task it fulfills.
- II. Find the neurophysiological realizers of that task
- III. Formulate a theory (in neurophysiological terms) of how these realizers carries out this causal task.

The crucial step in this reduction is the first one. Because what is being done is a form of “reduction as translation” in which a great loss of meaning is involved, an elimination of the conscious aspects of the mental state. What then follows in steps 2 and 3 is a causal reductive account - not of a conscious state, but of the behavior that the conscious state has been (re)defined as. Kim is clear that his

¹³³ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 100

¹³⁴ Kim, Jaegwon *Physicalism, or Something near Enough* (2005) p. 105

¹³⁵ Kim, Jaegwon *Physicalism, or Something near enough* (2005) p. 101-102; p. 164-165

¹³⁶ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) pp. 108

reduction involves a redefinition of the term to be reduced: “we must first define, or redefine it in terms of the causal task the property is to perform”.¹³⁷ The problem is that whereas this was done with integrity of meaning in the case of reproduction, there is a great loss of meaning in the functional reduction of mental states. Because in this “redefinition” the conscious, subjective aspects are left out.

The crucial step is the first one: mentality is defined in non-mental terms. And when we are trying to find the physical realizers of the behavior and not of the conscious state, and these realizers are in the neurophysiological domain, it is even more clear that any form of mental causation is out of the question (unless neurophysiological phenomena are conscious).

Let us say that we want to reduce the mental state I am in if I believe that there is an election today. The causal task this belief fulfills might perhaps be defined with reference to the behavior “going to the polling station and voting”. Step two then involves finding the physical realizers of that *behavior* with which we have identified the belief. And step three involves formulating a theory in physiological terms how these realizers makes this behavior happen. Nowhere is a mental event causally efficacious in virtue of mental properties.

Believing that there is an election today is defined in terms of behavior not in terms of propositional content and attitudes, and that behavior is performed by some neurophysiological realizers. Kim writes: “even though a complete analysis of belief is not in and perhaps never will be, we do not think there is ultimately anything beyond causal work vis-à-vis observable behavior that is involved belief”.¹³⁸ So the physical realizers that we try to find in step 2 and 3 are not realizers of a conscious, subjective mental experience, it is of the behavior “going to the polling station”.

If we instead in the first step preserved the mental, conscious state, and in the second step asked for its realizers, we would not get an ontological reduction of the mental to the physical, only a causal reduction. And with that the non-reductive materialists would be very happy. But then we could ask (again) whether the conscious state, realized by neurophysiology, was causally active as conscious or if all causality took place at the preconscious level, which is the question of epiphenomenalism and causal closure of the physical domain which Kim tries to solve, by defining mentality in non-mental terms.

All cognitive and intentional states can be defined in non-mental terms according to Kim: “as far as intentional states are concerned, we are within the domain of behavior and the physical mechanisms involved in their production; they do not take outside this domain.”¹³⁹ Intentional states are defined as the causal role they fulfill. But in order to fulfill a causal role, you must first exist. But it is no longer the intentional, conscious state *qua* conscious that fulfills the causal role, it is the physical properties P that performs the causal task. As Kim writes:

¹³⁷ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 101

¹³⁸ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 167

¹³⁹ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 168

“Having M = *def.* having some property other P (in the reduction base domain) such that P performs causal task C.”¹⁴⁰

Given Kim’s definition of mental causation, namely “that an event, *in virtue of its mental property*, causes another event to have a certain physical property”¹⁴¹ this simply does not make room for mental causation. It is on the one hand rather mindboggling that Kim presents this as a way of *saving* mental causation, since it excludes the defining character of mental states, namely its conscious subjective aspects, but it is on the other hand quite clear why Kim could not afford his definition to refer to *conscious* states.

If he did, the second step would presumably involve finding the realizers of this conscious state. But then the question could be put: does this conscious state when realized, that is, *qua* conscious, exert any causality, or is it only the physical realizers of the conscious state that exerts causal powers, by producing the conscious state and the behavior? If he answers this in the affirmative, yes, pain *qua* conscious exerts causal power, he would be left with the problem he began with, namely how something mental could affect something physical, given the causal closure of the physical domain. But if he answers in the negative, he is left with epiphenomenalism which he wanted to escape. Kim argues that:

“There is an honest difference between elimination and conservative reduction. Phlogiston was eliminated, not reduced; temperature and heat were reduced not eliminated.”¹⁴²

This is quite true. But the “reduction” of heat and temperature consisted in describing their underlying causes, a causal reduction. But in Kim’s functional reduction he first *eliminates* consciousness and then describes the underlying (physiological) causes of *behavior*. True, there is an honest difference between denying there being any behavior, and finding the neurophysiological realizers of behavior. But there is no difference between eliminating the conscious aspects of a mental state and defining it in non-mental terms. And it is quite remarkable that Kim, given his insistence on saving mental causation, defines that which he wants to save in non-mental terms. He affirms mental causation only in name and not in content.

Kim’s criticism of Davidson (and Loewer) was that their view of mental causation did not save human agency: “we want agents, in virtue of the beliefs and desires and intentions they hold, to cause their limbs to move in appropriate ways and thereby produce changes in their physical surroundings”.¹⁴³ One could say, perhaps, that if one *defines* “having a mental state” as the same, as “performing a function” then it does give room for mental causation (since mental causation does not involve conscious mental states). It would be a form of reduction through translation, but a form in which you have eliminated exactly that which Kim said he wanted to save. But if Kim in fact is

¹⁴⁰ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 101

¹⁴¹ Kim, Jaegwon “The Myth of Nonreductive Materialism” (1989) p. 43 Italics in original

¹⁴² Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 160

¹⁴³ Kim, Jaegwon “Responses” in *Philosophy and Phenomenological Research* Vol. 65:3 (2007) p. 675 <http://onlinelibrary.wiley.com/doi/10.1111/j.1933-1592.2002.tb00231.x/pdf> (retrieved: 2014-08-10)

perfectly fine with understanding mental causation as a form of non-conscious causation, neither his critique of Davidson nor Loewer makes sense. Indeed, if “believing x” should be understood as behaving in certain ways, and the cause of that behavior might very well be merely physical properties (which is what Kim’s definition implies) then Davidson’s solution is perfectly acceptable.

Kim’s argument for the definability of intentions and desires in terms of behavior rests among other things on his claim that would be “incoherent” to not assign beliefs and desires to creatures that are behaviorally indistinguishable from beings like us.¹⁴⁴ But the reason it would be incoherent, is not an analytical reason, in the way that it would be incoherent to deny that a bachelor is unmarried. It would be incoherent because the reasons we have for affirming that someone is conscious would be present in both cases - we are talking about the epistemological ground for affirming something to be the case, not what the case that we affirm is. Kim is here confusing epistemology and ontology (as functionalists tend to do). Kim argues, not that we are affirming mentality on their part because their behavior gives us *reasons for believing* that they are conscious and intentional, but because we – it is assumed - can *define* mental states in virtue of behavior, that “we cannot avoid thinking of intentional/cognitive states, like thought, belief, and desire as supervenient on behavior and other observable physical facts”.¹⁴⁵

I surely can avoid that, but let us assume that Kim cannot. The trouble (for Kim) is that if beliefs and desires really were supervenient on behavior, they would not – given Kim’s understanding of supervenience - be the causes of that behavior and therefore there would be no mental causation in these cases. According to Kim’s understanding of supervenience, supervenient phenomena are totally *dependent on and determined* by their subvenient base.¹⁴⁶ In that case intentions are not the causes of behavior, instead intentions are determined by the behavior.

It is fascinating to see that Kim uses exactly the same arguments against eliminating the conscious aspects of qualia from the definition of qualia, as one could use regarding other mental states: “Pain may be associated with certain causal tasks, but these tasks do not define or constitute pain.”¹⁴⁷ The defining character of pain is instead its conscious, subjective, experiential character: “as long as it hurts its pain”¹⁴⁸ regardless of whether you behave or not in certain ways. And since Kim agrees that it is impossible to make this redefinition of pain in non-mental terms, it cannot be accommodated within a physicalist worldview: “They stay outside the physical domain, but they make no causal difference and we won’t miss them.”¹⁴⁹

Pain is among these properties that are irreducible and therefore make no causal difference, according to Kim. So in the end there are mental phenomena which can be defined as something non-

¹⁴⁴ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 165

¹⁴⁵ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 165

¹⁴⁶ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 34

¹⁴⁷ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 169

¹⁴⁸ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 169

¹⁴⁹ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 173

mental, and they are causally efficacious, but of course not in virtue of any mental property (as that is generally understood) and there are mental phenomena which cannot be defined in non-mental terms and they are causally impotent. In neither case you get mental causation as Kim has described it:

Why do we care about mental causation? Because, first and foremost, we care about human agency. To save agency, however, we *need* the productive concept of causation; we want agents, in virtue of the beliefs and desires and intentions they hold, to cause their limbs to move in appropriate ways and thereby produce changes in their physical surroundings.¹⁵⁰

To me it is inexplicable how Kim can be so clearheaded in his critique of others accounts of mental causation as being inadequate and then provide this account. In any case, it seems as if there is a lot of trouble with mental causation *qua* mental in a physicalist ontology. So why not give it up? The next chapter is intended to show that this is not a viable strategy for anyone wanting to make a rational argument for anything at all - including physicalism.

Chapter 3. The Evolutionary Argument against Naturalism

3.1 Alvin Plantinga

Alvin Plantinga was born in 1932 in Michigan, as the son of Dutch immigrants. He received his doctoral degree in philosophy from Yale in 1958, and then taught as professor of philosophy at Wayne State University. In 1963 Plantinga returned to his *Alma Mater* Calvin College and taught there for nineteen years, before taking up the position as the John A. O'Brien Professor of philosophy at Notre Dame in 1982. His inaugural lecture was titled *Advice to Christian Philosophers* and it received a lot of attention and has been the subject of much discussion. In this lecture Plantinga argued that Christian philosophers should not necessarily be concerned with the problems that are in vogue in contemporary philosophy, but should consider what kind of philosophical problems that are of most interest for them as Christians, that is as philosophers that have some Christian commitments.¹⁵¹ Plantinga's own philosophical work has mainly been in metaphysics, epistemology and philosophy of religion.¹⁵² *The Evolutionary Argument against Naturalism* can be seen as combining these three interests.

3.2 The argument – introductory remarks

The Evolutionary Argument against Naturalism (EAAN) has received a great deal of interest, but although the issue of the compatibility between physicalism and mental causation is very central to the argument it has not received much attention in the philosophy of mind-literature.¹⁵³

¹⁵⁰ Kim, Jaegwon Kim, Jaegwon "Responses" (2002) p. 675

¹⁵¹ Plantinga, Alvin "Advice to Christian Philosophers" *Faith & Philosophy* Vol. 1 Issue 3, July 1984, available at: http://www.faithandphilosophy.com/article_advice.php (retrieved:2014-08-10)

¹⁵² <http://philosophy.nd.edu/people/alvin-plantinga/> (retrieved: 2014-06-02)

¹⁵³ There are some exceptions: Plantinga has been discussed in a couple of articles in *Journal for Consciousness studies*: "Evolution and Epiphenomenalism" by William S Robinson in *Journal for Consciousness Studies*, 14

It first appeared as an article by Plantinga in the publication *Logos* in 1991 and then in Plantinga's second book in his trilogy on epistemology, called *Warrant and Proper Function* in 1993.¹⁵⁴ It has later been published in several contexts, such as *Warranted Christian Belief* (2000) *Naturalism Defeated?* (2002) *Knowledge of God* (2008) and in the entry *Religion and Science* in *Stanford Encyclopedia of Philosophy* (2010).¹⁵⁵ The, as present, latest (and, as Plantinga says, "hopefully final") version was published as chapter ten in *Where the Conflict Really Lies: Science, Religion, and Naturalism* (2011). Although the presentations differs somewhat, the basic argument is the same. I will base my presentation from the last version (and in so far as I borrow anything from earlier versions it will not be in a respect that changes anything from the logic in the argumentation in the latest version, but only because I think the same point is stated a bit clearer in an earlier presentation).

3.3 The argument – some conceptual clarifications

There are a few central concepts in Plantinga's argument that I thought I should try to define.¹⁵⁶ The first concept is *naturalism*. Naturalism is a philosophical position that denies the existence of God or any other supernatural entity.¹⁵⁷ Plantinga calls naturalism a quasi-religion since it fulfills some of the functions that religions typically do, in the sense that it tries to answer fundamental questions about the nature of reality.¹⁵⁸ It would perhaps be better to call it a worldview.¹⁵⁹

Another thing to note is that for Plantinga naturalism, in this context, should be understood as including *materialism*.¹⁶⁰ Plantinga does not think that the arguments for materialism are conclusive for a naturalist, but he finds them to be strong and notes that most naturalists find them compelling, and for that reason he thinks it reasonable to think of naturalism as including materialism.¹⁶¹ For Plantinga, materialism means that humans are their bodies, or some part of their body, such as their brain, or some part of their brain, such as the left hemisphere.¹⁶² He distinguishes between reductive and non-reductive materialism but does not consider emergence theory.¹⁶³

Cognitive faculties are those faculties by which we form beliefs about the world, and among them Plantinga lists sense perception, memory, introspection (or self-reflection) and "a priori

(11) (2007) available at <http://www.public.iastate.edu/~wsrob/EvoEpi.pdf> and "What does the evidence tell us about the biological value of consciousness?" by Brian Earl *Journal of*

Consciousness Studies, 15, No 7, 2008 available at: http://www.imprint.co.uk/pdf/15_7-Earl.pdf

¹⁵⁴ Plantinga, Alvin *Warrant and Proper Function* (1993) chapter 12

¹⁵⁵ Plantinga, Alvin "Religion and Science" in *Stanford Encyclopedia of Philosophy* <http://plato.stanford.edu/entries/religion-science/> (retrieved: 2014-06-09)

¹⁵⁶ That is, I will try to clarify in what sense Plantinga uses them

¹⁵⁷ Plantinga, Alvin in Beilby, James (ed) *Naturalism Defeated?* (2002) p. 1

¹⁵⁸ Plantinga, Alvin *Where the Conflict Really Lies: Science, Religion, and Naturalism* (2011) p. 311

¹⁵⁹ As Carl Reinhold Bräkenhielm does in his book *Filosofiska livsåskådningar* (2008) p. 28

¹⁶⁰ Plantinga, Alvin *Where the Conflict Really Lies: Science, Religion, And Naturalism* (2011) p. 318-320

¹⁶¹ Plantinga, Alvin *Where the Conflict Really Lies: Science, Religion, And Naturalism* (2011) p. 320

¹⁶² Plantinga, Alvin *Where the Conflict Really Lies: Science, Religion, And Naturalism* (2011) p. 318

¹⁶³ Plantinga, Alvin *Where the Conflict Really Lies: Science, Religion, And Naturalism* (2011) p. 322-325

intuition” by which Plantinga means our ability to see logical connections between propositions (logical reasoning seems to be an appropriate alternative term).¹⁶⁴

The notions *guided* and *unguided evolution* are very central to Plantinga’s argument. As his use of these terms is somewhat ambiguous, I will interpret them in the way I think is most charitable. In the book *Science and Religion: Are They Compatible?* (2011) Plantinga defines unguided evolution as “the claim that no personal agent, not even God, has guided, planned, intended, directed, orchestrated, or shaped this whole process”.¹⁶⁵ At other times Plantinga writes as if guided evolution demands more than just intending some outcomes of and sustaining the processes, that it also demands that God has been an additional factor in the evolutionary processes by, for instance, causing “the right mutations to arise at the right time”.¹⁶⁶ I believe it is possible to interpret Plantinga in the more strict sense according to which unguided means “unplanned” without specifying how God must bring his intentions about. Plantinga allows that God might have chosen the right initial conditions rather than interacting with the process in time.¹⁶⁷

3.3.1 The expression “reliable” in reliable cognitive faculties

Since my discussion of the concept “reliable cognitive faculties” will be quite long, I thought it best to place it under a separate heading. In the first book in his trilogy on epistemology, *Warrant: the Current Debate* (1993) Plantinga analyzed the contemporary notions of warrant, that is, what it is that distinguishes knowledge from merely true belief.¹⁶⁸ One very influential view on warrant is justification: “[k]nowledge is justified, true belief: so we thought from time immemorial”¹⁶⁹ as Plantinga says. Plantinga introduces the problem for this view (in *Warrant and Proper Function*) by analyzing how the so-called Gettier problems seem to show that even justified true beliefs sometimes can fail to be knowledge in the proper sense.¹⁷⁰

Plantinga argues that justification is a matter of fulfilling your (alleged) epistemic duties: you are justified in believing p if have good reasons for believing p.¹⁷¹ You are epistemologically responsible or justified in believing p, if and only if, you have good reasons for believing p. Then your belief that p has warrant and will, if your belief is true, constitute knowledge. This view of warrant is called “internalist”, since it presupposes that the reasons for believing p are internally or cognitively accessible for you (if your belief is to be warranted).¹⁷² Plantinga thinks that Gettier problems show that the justification approach to warrant cannot be true, since they show that even justified true belief sometimes fails to constitute knowledge.

¹⁶⁴ Plantinga, Alvin *Where the Conflict Really Lies: Science, Religion, and Naturalism* (2011) pp. 311

¹⁶⁵ Plantinga, Alvin & Dennet, Daniel C. *Science and Religion: Are They Compatible?* (2011) p. 4

¹⁶⁶ Plantinga, Alvin *Where the Conflict Really Lies: Science, Religion, and Naturalism* (2011) p. 11

¹⁶⁷ Plantinga, Alvin *Where the Conflict Really Lies: Science, Religion, and Naturalism* (2011) p. 39, footnote 14.

¹⁶⁸ Plantinga, Alvin *Warrant and Proper Function* (1993) p. 36

¹⁶⁹ Plantinga, Alvin *Warrant and Proper Function* (1993) p. 31

¹⁷⁰ Plantinga, Alvin *Warrant and Proper Function* (1993) p. 36

¹⁷¹ Plantinga, Alvin *Warrant and Proper Function* (1993) p. vi

¹⁷² Plantinga, Alvin *Warrant and Proper Function* (1993) p. v

All Gettier problems have the same structure: you believe a proposition p, p is true, and you have good reasons for believing in p (so you are justified in believing p). According to one version of this problem, you wonder what time it is. Your (normally reliable) clock says that it is 12 o'clock. Since you are a rational epistemic subject, you then form the belief that it is 12 o'clock, which indeed turns out to be the case. According to this definition of warrant, you would then have known that it was 12 o'clock, since you had a justified, true belief that it was 12 o'clock. But, alas, the clock was actually not functioning, when you went to bed at 11:30 the clock stopped at 12 and when you went into the kitchen the next day, at 12 o'clock, it was just out of luck that it showed the correct time (that you happened to look at it at the time it was stuck on).

Did you actually *know* that it was 12 o'clock? Most would say no, and if you did not know it, then there is some problem with this understanding of warrant (because warrant is what is supposed to distinguish knowledge from mere true belief).

Plantinga's view on warrant is externalist. Plantinga does not think that a subject must know why she believes a certain proposition in order for that belief to enjoy warrant (and constitute knowledge). Plantinga instead argues that warrant should be seen as a question of whether the belief is produced by reliable cognitive faculties, that is, faculties that have a high objective probability of producing true beliefs.¹⁷³ But, as more than one thinker has observed, many of our beliefs have turned out to be false, so maybe our cognitive faculties are not reliable at all (in the sense of having a high probability of producing true beliefs)? Thus, if Plantinga's understanding of warrant is correct, it would make it impossible for us to ever achieve warrant. And of more relevance for this essay, the reliability, or at least the belief that our cognitive faculties are reliable, is a starting point for his *Evolutionary Argument against Naturalism*. As Plantinga says:

EAAN begins from certain doubts about the *reliability* of our cognitive faculties, where, roughly, a cognitive faculty – memory, perception, reason – is reliable if the great bulk of its deliverances are true.¹⁷⁴

Plantinga wants to argue that this reliability of our cognitive faculties' demand that they have been brought about through an evolutionary process that is superintended and orchestrated by God.¹⁷⁵ If our cognitive faculties are not reliable, then his argument for the need for that orchestration fails. And some think that this observation (that many beliefs are false) is devastating for Plantinga's argument:

"Plantinga's argument seems so weak as to barely justify further discussion. As Robinson notes, beliefs are often incorrect. He writes (p. 38) that '...beliefs in phlogiston, contingent identities, and communications from the dead come to mind as examples.' And we can think of many more everyday examples. Robinson asks (p. 38), 'How can these facts [about obviously false beliefs] be made compatible with the view that evolution selects for reliable processes of belief formation?' – which would seem to be the end for Plantinga's argument against epiphenomenalism."¹⁷⁶

¹⁷³ Plantinga, Alvin *Warrant and Proper Function* (1993) p. 46

¹⁷⁴ Plantinga, Alvin in Beilby, James *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument against Naturalism* (2004) p. 2

¹⁷⁵ Plantinga, Alvin in Beilby, James *Naturalism Defeated?* (2002) p. 3

¹⁷⁶ Earl, Brian "What does the Evidence Tell Us about the Biological Value of Consciousness?" *Journal of Consciousness Studies*, 15, No 7, 2008, p. 90 http://www.imprint.co.uk/pdf/15_7-Earl.pdf

Indeed, many beliefs are false, as this observant philosopher points out. He does not seem bothered by the implication of the position he argues for though (namely that our cognitive faculties are not reliable) and that might indicate that maybe he is not as observant as he could be. If it is indeed the case that our cognitive faculties are fundamentally unreliable, since epiphenomenalism is true and the content of our beliefs is irrelevant for our behavior, then that poses a problem not only for Plantinga's evolutionary argument against naturalism but for anyone interested in making a rational argument for any position at all. The belief that our cognitive faculties are fundamentally unreliable – which would be natural to assume if the content of our belief is irrelevant from an evolutionary standpoint and they have been formed through the processes of evolution - is, as Plantinga says “an undercutting defeater”¹⁷⁷ for any belief at all, in the sense that it takes away any rational grounds for holding any belief (including the belief that our cognitive faculties are unreliable).¹⁷⁸

Fortunately it is not the end for Plantinga's argument, because although Plantinga often imprecisely formulates the criteria of reliability of our cognitive faculties as implying that “the great bulk of its deliverances are true” his understanding of reliability is a bit more complex than that. Plantinga does take into account that reliability is not sufficient although it is necessary for forming warranted, true beliefs.

The propositional content of our beliefs (and so the truth or falsehood of a belief) must of course depend on more than only the reliability of the information-processing of our cognitive faculties, it must also depend a lot on what kind of information one (so to speak) feeds those cognitive faculties with. If I tell you a lie, and you believe me, it is not necessarily because your cognitive faculties are “unreliable” that you form a false belief.¹⁷⁹ If one analyzes how Plantinga understands “warrant” it becomes clear that Plantinga does take the cognitive environment (that is, the information which we encounter and process with our cognitive faculties) into account. A belief has warrant, according to Plantinga, if:

- a) “It has been produced in me by cognitive faculties that are working properly (functioning as they ought to, subject to no cognitive dysfunction) in a cognitive environment that is appropriate for my kinds of cognitive faculties)
- b) The segment of the design plan governing the production of that belief is aimed at the production of true beliefs, and
- c) There is a high statistical probability that a belief produced under those conditions will be true.”¹⁸⁰

The expression “the design plan” only refers to the observation that our cognitive faculties have functions, and that at least in some cases the principal function of them is to form true beliefs about the world. The third criteria is referring to the reliability of our cognitive faculties, let us call it the

¹⁷⁷ Plantinga, Alvin *Warrant and Proper Function* (1993) p. 230

¹⁷⁸ Which takes you into what Plantinga calls “one of those nasty dialectical loops” (1993) p. 234

¹⁷⁹ If I consistently lie to you and you do not develop any skepticism towards me, *then* one could say that your cognitive faculties are not reliable. But since people in general are not pathological liars, the principle of credulity as an initial attitude is a mark of properly functioning cognitive faculties, Plantinga would say. (See: *Warrant and Proper Function* (1993) pp. 33-34

¹⁸⁰ Plantinga, Alvin *Warrant and Proper Function* (1993) p. 46

“reliability criteria” and that criteria says, not that all our beliefs will have a high statistical probability of being true, but only that there will be a high statistical probability that a belief produced under *those conditions* will be true. One of those conditions is that the cognitive environment is “appropriate” and the demands on an appropriate cognitive environment turn out to be rather high. As Plantinga says in his discussion of Gettier problems in relationship to his own account of warrant:

So the first thing to see about the Gettier situations is that the true beliefs in these situations are true by accident, not by virtue of the proper function of the faculties or belief-producing mechanisms involved. And the second thing to see is that in the typical Gettier case, the locus of the cognitive glitch is in the cognitive environment: the latter is in some small way misleading.¹⁸¹

Plantinga does not think that Gettier problems destroy his conception of warrant, since Gettier problems involves a situation in which the environment is “in some small way misleading” – and if it is, it is not what Plantinga calls “a cognitive environment appropriate for my kinds of cognitive faculties”. That’s why his account of warrant isn’t vulnerable to Gettier problems as the internalist accounts of warrant are: the cognitive environment is in those cases not “appropriate”. Therefore it is only when the cognitive environment provides me with true and relevant information that my cognitive faculties should, according to Plantinga, be “producing, for the most part, true beliefs” if they are reliable. Many, probably most, false beliefs have been formed based on information that is (at least) “in some small way misleading”. Misleading information might also be information that leaves out something that is very relevant or central. If I pretend to tell you how much I earn each month but only tell you how much I earn from one of my ten employers, your belief will quite likely be false, because my information, though true, was not complete, with regard to relevant information.

And this turns out then to be an understanding of reliability that I believe most would agree that our cognitive faculties live up to: given true and (all) relevant information, our cognitive faculties are reliable, in the sense that they will produce, for the most part, true beliefs, when they are not subject to cognitive dysfunction. And the question that Plantinga asks should therefore be understood as: what is the probability that a naturalistic evolutionary process has produced cognitive faculties that are “reliable” in the sense of producing, when they are functioning properly, for the most part true beliefs, if given true and (all) relevant information?¹⁸²

Although it is at times unclear what is meant by unguided evolution, it is clear, however that, according to Plantinga, there is no evolutionary selection pressure towards reliable cognitive faculties. If there were, an unguided evolutionary process might be thought, with some fair degree of probability, to lead to beings with reliable cognitive faculties.

¹⁸¹ Plantinga, Alvin *Warrant and Proper Function* (1993) p. 35

¹⁸² If this is *not* what Plantinga means, this is a position that is consistent with his broader views on epistemology and which is much more reasonable than the so easily criticized idea that my cognitive faculties will form true beliefs, regardless of the quality of the information that they have access to. Working from the premise that one should build up a position worth attacking, I will interpret Plantinga in this way (which I also think is the correct interpretation of Plantinga).

When one analyzes why Plantinga does not think there is any reason to believe that there has been a selective pressure towards reliable cognitive faculties, it seems as if the solution (“guided evolution”) is not enough to get what most theists would want in terms of a philosophical anthropology. (Not only that we can know, but also that we can act out of this knowledge.) It also seems as if this solution (cognitive faculties, produced by a guided evolutionary process) is not available for Plantinga himself, given his own views on the mind-body problem.¹⁸³

3.3.2 “Undercutting” and “rebutting” defeaters

The notion of a defeater is central to Plantinga’s argument. A defeater is a proposition that, if true, lowers the probability of or support for the truth of another proposition. The difference between a rebutting and an undercutting defeater is that a rebutting defeater is evidence against a proposition, whereas an undercutting defeater takes away the support for believing the proposition. Let us say that a lot of trustworthy people tell me that Jim is a constant liar. This would constitute a rebutting defeater against the proposition “Jim is trustworthy.” Let us also say that Jim said that he knows that Lisa and Stephen will get married. The information that Jim is lying a lot, my knowledge about that, constitutes a rebutting defeater to the proposition that Jim actually knows that Lisa and Stephen will get married, and an undercutting defeater to any belief based upon Jim’s testimony, that Lisa and Stephen will get married, but not a rebutting defeater against the proposition that Lisa and Stephen will get married. I have neither more nor less evidence for the proposition that Lisa and Stephen will get married, but I do have evidence against the belief that Jim knows whether they will get married or not.

3.4.1 The structure of the argument

Before analyzing how Plantinga argues for the specific premises in the EAAN I thought I would just give a short overview of the structure of the argument:

Premise 1: Naturalism means that no God or no being like God exists and therefore no one has superintended or orchestrated the evolutionary process to produce beings with reliable cognitive faculties.

Premise 2: Naturalism includes materialism.

Premise 3: Given materialism, propositional content and qualia are not causally efficacious in virtue of their propositional content or conscious aspects. Only the neurophysiological properties of conscious events are causally efficacious.

Premise 4: Evolution is not interested (so to speak) in what we feel or think, unless that affects our behavior.

¹⁸³ Plantinga is a dualist, and he has argued for this position in several publications, including *Knowledge of God* (2008) and in the article “Materialism and Christian Belief” in van Inwagen, Peter & Zimmerman, Dean (ed) *Persons: Human and Divine* (2007)

Premise 5: Given premise 3, the semantic content of our beliefs and the conscious aspects of our feelings are irrelevant from an evolutionary standpoint.

Conclusion 1: Given naturalism, there are no selection pressures towards cognitive faculties that tend to produce true rather than false beliefs, and we have therefore no reason to trust the propositional content of our beliefs, if we believe they have been produced by such a process.

Conclusion 2: Conclusion 1 constitutes an undercutting defeater for any belief produced by these cognitive faculties, including the belief that naturalism is true, which means that naturalism cannot be rationally held.

3.4.2 Analyzing the argument

I will now, as carefully as I can, present Plantinga's latest version of this argument, which is in *Where the Conflict Really Lies* (2011).¹⁸⁴ Plantinga begins his presentation with making the distinction between guided and unguided evolution. Plantinga says that the claim that evolution is "unguided" is a theological add-on and not a part of the scientific theory of evolution as such – and that is certainly true if we understand "unguided" in the broader sense.¹⁸⁵ Secondly he presents what is going to be the thesis for which he aims to argue in this chapter, that there is a conflict between evolution (or more specifically, the aspect of evolutionary theory which is the belief that our cognitive faculties have evolved by the evolutionary processes) and *naturalism*, which implies the claim that evolution is unguided – "not guided or orchestrated by anyone" – for instance by choosing the initial conditions that might lead to beings with reliable cognitive faculties.¹⁸⁶ The conflict is not such that it cannot be *true* that our cognitive faculties have evolved through a naturalistic evolutionary process, it is such that it cannot be rationally believed.¹⁸⁷ As Plantinga says:

By way of analogy: I cannot sensibly believe that there aren't any beliefs, or that no one has true beliefs, or that my beliefs are all false.¹⁸⁸

Then Plantinga presents where the conflict is supposed to lie, between naturalism and evolution: with regard to the reliability of our cognitive faculties.¹⁸⁹ Reliability is to be understood, as explained earlier, as the probability that our cognitive faculties, when they are not malfunctioning and the cognitive environment is appropriate for them, will produce, for the most part, true beliefs.¹⁹⁰ Plantinga argues that, if our cognitive faculties have been formed by only those forces that current evolutionary theory talks about (mainly natural selection and random genetic mutation) then you

¹⁸⁴ Plantinga, Alvin *Where the Conflict Really Lies* (2011) chapter 10

¹⁸⁵ Plantinga, Alvin *Where the Conflict Really Lies* (2011) p. 308-309

¹⁸⁶ Plantinga, Alvin *Where the Conflict Really Lies* (2011) p. 308-309

¹⁸⁷ Plantinga, Alvin *Where the Conflict Really Lies* (2011) p. 310

¹⁸⁸ Plantinga, Alvin *Where the Conflict Really Lies* (2011) p. 310

¹⁸⁹ Plantinga, Alvin *Where the Conflict Really Lies* (2011) p. 312

¹⁹⁰ Plantinga, Alvin *Where the Conflict Really Lies* (2011) p. 312

cannot “sensibly think that our cognitive faculties are for the most part reliable”.¹⁹¹ The reason is that from the evolutionary standpoint, the content (and so the truth or falsehood) of our beliefs is not relevant in and of themselves.¹⁹² Beliefs are, in themselves, invisible for evolution; the only way in which they might be “visible” is if they affect our behavior.¹⁹³ This observation leads to what Plantinga calls *Darwin’s doubt* - a doubt in the adequacy of evolution for producing reliable (in the aforementioned sense) cognitive faculties. The basic worry can be formulated in a formula, which Plantinga does: what is the probability (P) that our cognitive faculties are reliable (R), given that (|) they have been produced by the process of evolution (E), and naturalism is true (and therefore no one has orchestrated or superintended the process) (N)? The formula looks like this: What is P(R|N&E)? The expression “given that” refers to what is called “conditional probability”. It refers to the fact that the probability for something being the case often depends on whether other facts hold. We are not asking about the probability of reliability of our cognitive faculties, in general, but *given that* they have been formed by a naturalistic evolutionary process.

Since evolution is not interested in our beliefs *per se* but only in our beliefs insofar as they affect our behavior, we must ask how we shall understand the relationship between our beliefs and our behavior. As mentioned earlier, Plantinga observes that most naturalists are materialists. Although Plantinga does not think that a naturalist is logically bound to be a materialist, it is the position he thinks most cogent, given naturalism (as most naturalists also seem to think). One reason for naturalists to be materialists that Plantinga quotes is that naturalists are (almost without exception) evolutionists and since it seems impossible for an immaterial substance to “come to exist by the processes that evolutionary science posits”¹⁹⁴ they have problems with substance dualism. For the purposes of this argument, then, Plantinga sees naturalism as including materialism.¹⁹⁵

The versions of materialism that Plantinga considers are reductive and non-reductive materialism. In neither of these versions it seems as if that the propositional content of beliefs could be causally efficacious. I will argue that they both represent what Plantinga has called “semantic epiphenomenalism” – which, according to Plantinga, is a very common view within philosophy of mind.¹⁹⁶

The first step, for Plantinga, in making the claim that semantic epiphenomenalism follows from non-reductive materialism is to point out that a belief, from a non-reductive materialist point of view, will have two kind of properties: neurophysiological properties and propositional or semantic properties.¹⁹⁷ That is, each belief that I have, will be represented or stored in my brain through a certain neurophysiological configuration and these are the neurophysiological properties of my belief.

¹⁹¹ Plantinga, Alvin *Where the Conflict Really Lies* (2011) p. 313

¹⁹² Plantinga, Alvin *Where the Conflict Really Lies* (2011) p. 31

¹⁹³ Plantinga, Alvin *Where the Conflict Really Lies* (2011) p. 316

¹⁹⁴ Plantinga, Alvin *Where the Conflict Really Lies* (2011) p.310

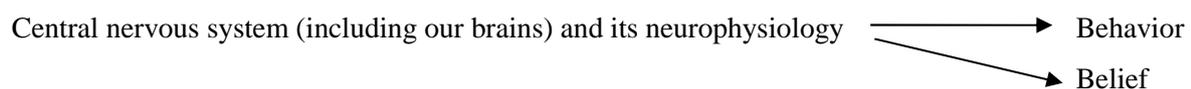
¹⁹⁵ Plantinga, Alvin *Where the Conflict Really Lies* (2011) p. 320

¹⁹⁶ Plantinga, Alvin in Beilby, James (ed.) *Naturalism Defeated?* (2002) p. 6

¹⁹⁷ Plantinga, Alvin *Where the Conflict Really Lies* (2011) p. 321

These aspects can be investigated by neuroscience. But there will also be semantic aspects to my belief, a content to my belief, the propositional content of my belief, or that which I believe.¹⁹⁸ An analogy can be made to written or spoken sentences, which have material aspects (the sound waves, or the ink on the paper that form the words) and semantic aspects (the propositional content of the sentence). It is, of course, in virtue of the propositional content that a belief can be true or false. Another example: if someone is singing in a very high tone, this might have the effect that a glass vase shatters. This happens, of course, in virtue of the material aspects of the singing. The song might also have a semantic content, which perhaps makes someone really happy, because it is a love song, dedicated to the listener. What is the relationship between the material and the semantic properties on the one hand, and the relationship between both of these and our behavior, according to non-reductive materialism?

According to non-reductive materialism semantic properties are not reducible to neurophysiological properties, but are determined by them, they *supervene* on them.¹⁹⁹ As Plantinga points out, supervenience implies a one-directional dependency: the mental is dependent on, and caused by, the physical, but not vice versa.²⁰⁰ We can therefore think of the central nervous system as fulfilling two functions: it produces behavior on the one hand, and it produces beliefs on the other. *How* the central nervous systems “produces” beliefs, we do not know. Given non-reductive materialism this is understood to be the result of complex material structures.²⁰¹ So somewhere in the history of evolution there arose an organism that acquired beliefs as a result of a more complex brain. But, according to supervenience (non-reductive materialism), the content of these beliefs is not causing anything, the content is merely supervening on the neurophysiology (which causes action). To put it in a graphic presentation:



To borrow a phrase from the literature on method in science, the CNS is an “independent variable”, behavior and beliefs are “dependent variables”. Since behavior is a dependent variable that is visible for natural selection there is a way in which evolution can give feed-back to the central nervous system, namely by selecting those that tend to produce adaptive behavior. But the same does not hold for the reliability (i.e. truth-conduciveness) of the belief-producing function of central nervous systems, since (given non-reductive materialism) beliefs do not affect my behavior in virtue of the content of these beliefs. This is the relationship between behavior and selection:

¹⁹⁸ Plantinga, Alvin *Where the Conflict Really Lies* (2011) p. 321

¹⁹⁹ Plantinga, Alvin (2011) p. 324

²⁰⁰ Which has led many to draw the same conclusion as Plantinga does, that on supervenience, mental properties are causally impotent. See, for instance Sven Walter’s article “Epiphenomenalism” at the *Internet Encyclopedia of Philosophy*, specifically the chapter on causal exclusion: <http://www.iep.utm.edu/epipheno/#SSH4b.iii> (retrieved: 2014-06-09)

²⁰¹ Plantinga, Alvin *Where the Conflict Really Lies: Science, Religion, and Naturalism* (2011) p. 325

Natural selection \longrightarrow behavior \longleftrightarrow behavior-producing aspect of the central nervous system

(The arrow from behavior to the CNS should be understood as an indirect continuation of the arrow from “natural selection” to behavior.)

There is a way for evolution to select different kinds of central nervous systems, by selecting those that produce adaptive behavior. But the same does not hold for reliable cognitive faculties (reliable in the sense of being conducive to true beliefs). Truth and falsehood are dependent on the propositional content of beliefs, and the propositional content of beliefs is (given non-reductive materialism) invisible for natural selection, since the content does not enter the causal chain leading to action. This means that natural selection would only be able to select for the aspects of the central nervous system that affects our behavior. The only way for the belief-producing aspects of the central nervous system to be selected for, is if our beliefs affect how we behave, in virtue of the semantic aspects of our beliefs, as shown below:

Natural selection \longrightarrow behavior \longleftrightarrow content of our beliefs \longleftrightarrow belief-producing aspect of the central nervous system.

(As before, the arrows pointing right are to be understood as an indirect continuation of the arrow from natural selection.)

If content is irrelevant for our behavior, what is the probability that a naturalistic evolutionary process has given us reliable cognitive faculties? It is of course very low. There is no reason at all why such a process would have led to reliable cognitive faculties. But if non-reductive materialism is true, our beliefs won't affect our behavior in virtue of their content, and therefore, if non-reductive materialism is true, $P(R|N\&E)$ is low. To sum up: according to non-reductive materialism the content of belief is dependent on the neurophysiological properties that causes both content and behavior. But whereas behavior is visible for natural selection, content is not. The important thing to note is that, as far as we know, there is no reason why the content caused by the neurophysiology that also causes behavior should in any way have any logical relationship to the behavior. As far as we know, the neurophysiology that makes my eyes blink might give rise to content such as “the moon is hot”.

What is the situation if reductive materialism is true? Plantinga understands reductive materialism to mean that mental properties “just are” physical properties.²⁰² What it means to say, as reductive materialists often do, that mental properties “just are” physical properties is not very clear, and I am doubtful that Plantinga's presentation of reductive materialism is sufficiently mindful of this ambiguity. Reductive materialism is also known as “identity theory” since the basic idea is that mental properties are identical to physical properties. Thus the in the literature oft-repeated expression “pain

²⁰² Plantinga, Alvin *Where the Conflict Really Lies* (2011) p. 333

just is c-fibers firing”.²⁰³ Plantinga’s treatment of reductive materialism is very sparse in *Where the Conflict Really Lies*. It is not possible to find any real difference between non-reductive and reductive materialism in the presentation in *Where the Conflict Really Lies* – on both versions it seems as if content is causally inert (which I believe is true) but some reductive materialists would probably object to that and a more careful treatment of reductive materialism seems to be appropriate. I will therefore complement Plantinga’s presentation in *Where the Conflict Really Lies* with two articles in which he deals with reductive materialism more fully.

In *Where the Conflict Really Lies* the difference between reductive and non-reductive materialism can be said to lie in how they answer the question of where content resides. Is it something that “supervenies on” the neurophysiology, or is it rather intrinsic to the neurology? Both positions agree that it is only the more complex neurophysiology that either give rise to content (through supervenience) or which also *is* content (on reductive materialism).²⁰⁴ One can say that for Plantinga, in *Where the conflict really lies*, this is a distinction without a difference, because in neither version it is, as Plantinga sees it, the content which is causally efficacious, *qua* content:

But now suppose materialism were true: then as we’ve seen, my belief will be a neural structure that has both NP [NeuroPhysiological] properties and also a propositional content. It is by virtue of the NP properties, however, not the content, that the belief causes what it does cause.²⁰⁵

In the article *Evolution, Epiphenomenalism, Reductionism* (2004) Plantinga argues against Jaegwon Kim’s claim that reductive materialism saves mental causation. According to Kim’s understanding, only physical phenomena can exert causal power.²⁰⁶ That is because Kim is committed to physicalism, and he takes physicalism to imply both the “ontological thesis” that “bits of matter and their aggregates exhaust the content of the world”²⁰⁷ and the principle of the causal closure of the physical domain.²⁰⁸ Kim’s proposed solution is to say that the mental is reducible to and identical with some physical state of affairs, which has causal powers. In his article *Content and Natural Selection* (2011) Plantinga seems to concede that reductive materialism means that the reductive materialist can escape semantic epiphenomenalism:

If content properties just are NP properties, there is no reason whatever for thinking content does not enter the causal chain leading to behavior. The specter of semantic epiphenomenalism is dispatched.²⁰⁹

This, however, is to go a bit fast. Exactly what do identity theorists mean when they say that “content” just are “NP properties”? The so-called causal argument for the identity theory is supposed to save mental causation, it goes like this: mental properties evidently have effects in the world, and since

²⁰³ Maslin, Keith T. *An Introduction to the Philosophy of Mind* (2007) p.

²⁰⁴ Plantinga, Alvin *Where the Conflict Really Lies: Science, Religion, and Naturalism* (2011)

²⁰⁵ Plantinga, Alvin *Where the Conflict Really Lies: Science, Religion, and Naturalism* (2011) p. 336

²⁰⁶ Kim, Jaegwon *Physicalism, or something near enough* (2005) p. 1

²⁰⁷ Kim, Jaegwon *Physicalism, or something near enough* (2005) p. 3

²⁰⁸ Kim, Jaegwon *Physicalism, or something near enough* (2005) p. 21

²⁰⁹ Plantinga, Alvin “Content and Natural Selection” in *Philosophy and Phenomenological Research* Vol. 83 No. 2 (September 2011) available at http://www.andrewmbailey.com/ap/Content_Natural_Selection.pdf (Retrieved: 2014-08-25)

according to *physicalism* every physical event (if it has a cause) has a *physical* cause, and since we (it is presupposed) should be physicalists, then we must understand mental properties to be physical.²¹⁰ But this will only work as a safeguarding of that which one wanted to save if one allows mental properties to exert causal powers *qua* (what is traditionally called) mental properties. For the causal argument to really work then, it would mean, for one thing, that physical objects *qua* physical could guide their behavior by virtue of rational reasoning. It is a radical expansion of the understanding of the concept “physical” and “physical causation” and makes this to some sort of pan-/localpsychism, according to which the material, *qua* material has (what is traditionally called) mental properties. It is a perfectly thinkable position, it just does not seem to be what the identity theorists really want.²¹¹

As Keith T. Maslin observes, although the identity theorists claim that “mental states = brain states” not all brain states are mental states (since there are processes in the brain which are not associated with conscious episodes, such as the processes in the brain stem that regulate our pulse):

These brain states will possess only physical properties. But those that are identical with conscious, mental states, must possess another layer of properties, namely mental ones, otherwise the identity of these brain processes with mental processes could not be asserted in the first place.²¹²

It follows that those brain states that are also mental states have more properties than those that are not mental states. And if we are to escape semantic epiphenomenalism then we must allow for some kind of causation, according to which these additional properties can exert causal powers in virtue of being those (what is traditionally called) mental properties.

And Plantinga does not understand reductive materialism in this panpsychistic way. How does he then think that reductive materialism escapes semantic epiphenomenalism, without still not leading to content causing behavior *qua* content? Well in the end that is not possible: if content *qua* content is causally inefficacious, then you have semantic epiphenomenalism. I think that Plantinga’s ambiguity reflects the ambiguity inherent in the mind/brain identity theory which in the end eliminates mental properties. I think Plantinga can be interpreted to mean that mental properties resides in material properties given reductive materialism and, since there is – as far as the reductive materialists has told us - no logical connection between the behavioral consequences of the neurophysiology and the content that a certain neurophysiological configuration constitutes, the neurophysiology *qua* content is irrelevant for behavior. True: if the neurophysiology were different, both the content and the behavior would be different. But it does not matter, since there is no logical relationship between the content that a certain neurophysiology is and the behavior it causes. So for instance, there is a neurophysiological state which causes me to scratch my chin – as far as we know (that is: for all the physicalist knows in virtue of his or her knowledge of the physical world) there is no reason why that neurophysiological state should also constitute a content which has a logical relation to the behavior it

²¹⁰ Kim, Jaegwon *Physicalism, or something near enough* (2010) p. 124-125

²¹¹ Maslin, Keith T. *An Introduction to the Philosophy of Mind* (2007) p. 90

²¹² Maslin, Keith T. *An Introduction to the Philosophy of Mind* (2007) p. 78

caused. Keith T. Maslin explains this by reference to the non-systematic relationship between colour and shape:

Every coloured object will have a shape and, equally, it is arguable, every item with a shape must possess a colour (...). But there will be no systematic relationship between the properties of colour and shape, such that an object can be, say, rectangular if, and only if, it is red, or that an object can be blue if, and only if, it is round.²¹³

Maslin applies this to the relationship between brain states and mental states in general, and the same goes, of course, to brain states (including their behavioral consequences) and the semantic content they constitute.

In the discussion in the article (“Content and Natural Selection”, 2011) Plantinga, although he allows that content is identical with some neural circuitry, still distinguishes between two aspects of the neurophysiological circuitry, the configuration *qua* behavior producing and the configuration *qua* content. Such a distinction is evident when Plantinga writes: “natural selection, in selecting for more adaptive belief-producing processes, won’t ordinarily select for more reliable belief-producing processes”.²¹⁴

It is the very same neural configuration that is adaptive (by causing adaptive behavior) and which is a certain content, but it is in virtue of causing adaptive behavior that natural selection selects it, and it is in virtue of its traditional neurophysiological properties that they cause behavior - and there is no logical connection between the semantic properties (or *a fortiori* the truth of the content) and the behavior-producing properties. Although, necessarily a certain neurophysiology will both be producing behavior and be a certain content, it is not in virtue of being that content that it is adaptive:

“The property *having q as content* is adaptive (we may suppose), but adaptive by virtue of the behavior it causes, not by virtue of its relation to that proposition *q*.”²¹⁵

It seems to me that this is, for all practical purposes, a form of semantic epiphenomenalism, since it pictures a situation in which it is *not* in virtue of the content *qua* content that an organism behaves as it behaves. Part of the confusion lies, I think, in that identity theorists are not very clear about what *they* mean when they say that mental properties are physical properties. Do they understand this as form of elimination, or do they understand it as a form of identification? And if it is an identification, it follows (since identities are symmetrical) not that mental properties no longer exist, but rather that physical systems are also mental.

If identity theorists argue, as they sometimes do, from the observation that mental properties have physical effects, and that epiphenomenalism therefore is untenable, to the position that the mental is physical, it does not hold to understand the identification of the mental to the physical in an eliminative sense (since then you have eliminated both the mental and its causal powers, and the

²¹³ Maslin, Keith T. *An Introduction to the Philosophy of Mind* (2007) p. 87

²¹⁴ Plantinga, Alvin “Content and Natural Selection” (2011) p. 440

²¹⁵ Plantinga, Alvin “Content and Natural Selection” (2011) p. 442

motivation for this move was to safeguard the causal powers of the mental, and if something is understood as having causal powers it must exist).

The only way in which the specific character of the semantic content could be relevant for behavior, given reductive materialism, would be if the neurophysiological properties that caused behavior (and constituted semantic content) caused behavior in virtue of the semantic content they constituted, which would imply some form of semantic causation that represents a radical expansion the present understanding of “physical causation”. Identity theorists are typically not interested in expanding the notion of the physical, they are interested in solving the mind-body problem without having to expand their ontology to include such things which are traditionally called mental, by *identifying* the (for their worldview) alien mental properties with the familiar physical properties. In line with that the identity-theorist Smart stated:

Some philosophers hold that though experiences are brain processes they nevertheless have fundamentally non-physical, psychical, properties, sometimes called ‘qualia’. Here I shall take the identity theory as denying the existence of such irreducible non-physical properties.²¹⁶

To sum up: either identity theory would lead to a form of pan-/localpsychism according to which the physical includes the mental (and thus making that which is traditionally called mental causation in principle possible) or also the semantic properties of beliefs are irrelevant for our behavior. I understand Plantinga to mean that both reductive and non-reductive materialism implies semantic epiphenomenalism.

A further possible interpretation of Plantinga should be mentioned. In earlier presentations²¹⁷ of the argument Plantinga listed four to five different ways in which content could be related to behavior. Among them was the possibility that content was causally efficacious, but still not leading to a selection pressure towards reliable cognitive faculties. Plantinga pointed out that what reliability demands is a selection towards truth-conducive cognitive faculties and what natural selection is after is not “truth” in and of itself, even when beliefs make a difference in virtue of their semantic content to our behavior, but rather beliefs that result in an adaptive behavior.²¹⁸ Is this Plantinga’s strategy towards reductive physicalism? That is, that he grants that semantic content *is* causally efficacious but argues that it is not enough for reliability since natural selection would select belief producing mechanisms that cause adaptive behavior, rather than those that produce true beliefs? I do not think so, because in *Where the Conflict Really Lies* which was published the same year as *Content and Natural Selection* (2011) Plantinga acknowledges that in general “true beliefs are more likely to be successful than false beliefs”²¹⁹ and this concession is probably the reason why this particular possibility is no longer

²¹⁶ Smart, JJC “The Mind/Brain Identity Theory” in *Stanford Encyclopedia of Philosophy* <http://plato.stanford.edu/entries/mind-identity/> (retrieved: 2014-07-31) Smart is not interested in expanding the notion of the physical, he rather denies that which we cannot currently describe as physical.

²¹⁷ Such as in *Warrant and Proper Function* (1993), and in Beilby, James (ed.) *Naturalism Defeated?* (2002)

²¹⁸ Plantinga, Alvin *Warrant and Proper Function* (1993)

²¹⁹ Plantinga, Alvin *Where the Conflict Really Lies* (2011) p. 335

considered in the latest version of the EAAN. It is still true though, that the naturalist needs both semantic content to be causally efficacious and that true semantic content in general leads to more adaptive behavior than false semantic content.²²⁰

However, if one understands reductive materialism as being compatible with semantic causality, then that implies a radical expansion of one's understanding of the nature of matter. And since that is seldom the case, both reductive and non-reductive materialism (as they are traditionally understood) will have the common effect that the propositional content of beliefs is irrelevant for our behavior, and therefore: $P(R|N\&E\&\text{reductive or non-reductive materialism})$ is low.

3.5 Why is this a problem?

Plantinga thinks that if there is no reason why a certain belief should be true (since it has been formed by cognitive faculties which have no interest in content) then the probability that that belief should be true could be estimated to 0.5, that it is equally likely that it is true or false.²²¹ I think that is a very high estimate, but even given such a high estimate, the probability that at least three quarters of 1000 beliefs would be true is very low: 10^{-58} (that is one zero followed by 58 zeroes). Why is this a problem?

The first premise in the EAAN is that $P(R|N\&E)$ is low. It is a problem because then we will, if we believe N&E, and see that $P(R|N\&E)$ is low have an undercutting defeater for any belief formed by those cognitive faculties that have been brought about by a naturalistic evolutionary process. So, if we believe N&E, and see that $P(R|N\&E)$ is low, we have a rebutting defeater for the reliability of our cognitive faculties.²²² But if we have a rebutting defeater for believing that our cognitive faculties are reliable, it also follows that we have an undercutting defeater for believing any belief formed by these cognitive faculties:

If you have a defeater for R, you will also have a defeater for any belief you take to be produced by your cognitive faculties, any belief that is a deliverance of you cognitive faculties. But *all* your beliefs, as I'm sure you have discovered, are produced by your cognitive faculties.²²³

That includes the belief that our cognitive faculties are formed by a naturalistic evolutionary process, or N&E for short. Therefore: if we believe N&E, and see that $P(R|N\&E)$ is low, we also get an undercutting defeater for believing N&E.²²⁴ N&E can still be true, but, as Plantinga said, we cannot

²²⁰ Sometimes our cognitive faculties seem to work to make us more optimistic than is realistic, but this is not the general case, and in general that does not seem to lead to adaptive behavior.

²²¹ Plantinga, Alvin *Where the Conflict Really Lies: Science, Religion, and Naturalism* (2011) p. 331

²²² Plantinga, Alvin *Where the Conflict Really Lies: Science, Religion, and Naturalism* (2011) p. 340

²²³ Plantinga, Alvin *Where the Conflict Really Lies: Science, Religion, and Naturalism* (2011) p. 344

²²⁴ The unique aspect of this defeater is that it cannot be defeated, because any evidence against the belief that our cognitive faculties are unreliable would itself be produced by those cognitive faculties, and therefore be subjected to the same undercutting defeater as any other belief. *Where the Conflict Really Lies:* (2011) p. 345 It seems that the only way to escape it is to abandon the reason for the defeater, that is N&E.

rationaly believe it. We have a rebutting defeater for the proposition that my cognitive faculties are reliable, and that constitutes an undercutting defeater for any proposition formed by these faculties.²²⁵

3.6 Evaluation of the problem for the materialist

I believe that Plantinga has clearly shown that if the content of our beliefs is irrelevant for our behavior and our cognitive faculties have been formed by a naturalistic evolutionary process, then P(R) is low, in which case we have an undercutting defeater for any belief produced by our cognitive faculties. On both reductive and non-reductive materialism (unless reductive materialism is interpreted in the panpsychistic direction) content is irrelevant for our behavior. But as Plantinga also points out, content is actually not irrelevant for our behavior. Plantinga agrees that it is obvious that in general “true beliefs are more likely to be successful than false beliefs”.²²⁶ But Plantinga points out that what we are asking is not “is content causally efficacious?” but: “can content be causally efficacious given materialism?” And although it might for many of us be obvious that content is causally efficacious, the important thing to notice is that there are a lot of advocates of reductive and non-reductive materialism out there, and Plantinga’s analysis has shown that their position is self-defeating.

3.7 The proposed solution

Plantinga points out, time and again, that he has no problems with the view that humans, including their cognitive faculties, have evolved, and that his argument is against “metaphysical naturalism” not evolution. What he is attacking is not the “E” in the conjunction “N&E” where “E” stands for the belief that we, including our cognitive faculties, have been brought about through the processes that “contemporary evolutionary science suggests”.²²⁷ It is the belief in an unguided (“naturalistic”) evolution that Plantinga argues leads to a self-defeating position. In *Warrant and Proper Function* (1993) Plantinga argued that a traditional Jew, Christian or Moslem has no reason for doubting the reliability of our cognitive faculties, even if he believes in evolution:

He may indeed endorse some form of evolution; but if he does, it will be a form of evolution guided and orchestrated by God. And *qua* traditional theist – *qua* Jewish, Moslem, or Christian theist – he believes that God is the premier knower and has created us human beings in his image, an important part of which involves his endowing them with a reflection of his powers as a knower.²²⁸

And in *Naturalism defeated?* Plantinga points out that he is not attacking the theory of evolution, and that he has no problems with the conjunction of evolution with theism. Evolutionary theory, Plantinga writes:

taken by itself (without the patina of philosophical naturalism that often accompanies expositions of it) is not as such in tension with the idea that God has created us and our cognitive faculties in such a way that the latter are reliable.²²⁹

²²⁵ As mentioned earlier, this gets you into what Plantinga calls “one of those nasty dialectical loops”. *Warrant and Proper Function* (1993) p. 234

²²⁶ Plantinga, Alvin *Where the Conflict Really Lies: Science, Religion, and Naturalism* (2011) p. 335

²²⁷ Plantinga, Alvin in Beilby, James *Naturalism Defeated?* (2002) p. 1-2

²²⁸ Plantinga, Alvin *Warrant and Proper Function* (1993) p. 236

²²⁹ Plantinga, Alvin in Beilby, James *Naturalism Defeated?* (2002) p. 2-3

The reason, for Plantinga, why a naturalistic evolutionary process cannot be said to, with a fair degree of probability, give rise to reliable cognitive faculties is that there is no selective pressure towards reliable cognitive faculties given a naturalistic evolution. The reason for this is that Plantinga takes naturalism to include materialism and since materialism does not allow content to be causally efficacious, the evolutionary process cannot, in itself, favor reliable (truth-conducive) cognitive faculties. Theism is able to solve this problem, since God could have:

guided and orchestrated the course of evolution, planned and directed it, in such a way as to achieve the ends he intends. Perhaps he causes the right mutations to arise at the right time; perhaps he preserves certain populations from extinction; perhaps he is active in many other ways.²³⁰

The solution to the problem which Plantinga has observed is not, for Plantinga, that beliefs must be causally efficacious (in virtue of their content) and that true beliefs in general are more adaptive than false beliefs, for us to be able to believe in the reliability of our cognitive faculties. It is that God has “guided, directed, orchestrated, or shaped”²³¹ the evolutionary process so that it would produce creatures with reliable cognitive faculties.

To sum up: according to Plantinga, the process of evolution does not as such favor reliable cognitive faculties. But the theist might believe that God has, in some way, acted as to make sure that the evolutionary process would produce beings with reliable cognitive faculties. Plantinga says that God might either choose the initial conditions that will, in time, lead to beings with the characteristics he intends (such as reliable cognitive faculties), or God might cause the right mutations to arise.²³² What Plantinga cannot afford is that there is a way (available for the naturalist) in which beliefs enter into the causal chain leading to behavior and that true beliefs in general are more adaptive than false beliefs. The reliability of our cognitive faculties must remain unexplained given naturalistic evolution, and it should be explainable given *guided* evolution. But guided evolution does not seem to be able to solve this problem, without at the same time undermining one of the central premises of the argument, namely that, given naturalism, semantic content cannot be causally efficacious.

3.8 Plantinga’s solution as inadequate for Plantinga (and other dualists of his kind)

As shown above Plantinga thinks there is a problem with unguided evolution: the resulting cognitive faculties are not likely to be reliable, which presents the naturalistic evolutionist with a defeater for any belief produced with those cognitive faculties. However, the proposed solution, *guided evolution*, does not seem to be available for Plantinga himself, given his own views on the mind-body problem. Plantinga is a dualist and a sharp critic of materialism regarding human beings.²³³ Plantinga points out three problems in particular for a materialist regarding our mental life:

²³⁰ Plantinga, Alvin *Where the Conflict Really Lies: Science, Religion, and Naturalism* (2011) p. 308

²³¹ Plantinga, Alvin *Where the Conflict Really Lies: Science, Religion, and Naturalism* (2011) p. 12

²³² Mutations: Plantinga, Alvin *Where the Conflict Really Lies: Science, Religion, and Naturalism* (2011) p. 12
initial conditions: *ibid* p. 39 note 14

²³³ See for instance, “Against Materialism” in *Faith and Philosophy* Vol. 23, No. 1 2006 & “Materialism and Christian Belief” in van Inwagen & Zimmerman (ed.) *Persons: Human and Divine* (2007)

- I. how can material structures acquire propositional content?
- II. how can something material be conscious,
- III. how can something material “think”?²³⁴

To that one could add the problem of mental causation. Although these questions are very difficult (which amounts to say that the mind-body problem for a materialist is a hard one) I do not intend to discuss them here. What I want to point out is that if it is *not* the case that our brains can give rise to conscious experience, harbor beliefs and think, the question arises: in what sense would a *guided* evolutionary process help us gain reliable cognitive faculties? Our cognitive faculties are those faculties, with the help of which we form beliefs. Among them Plantinga lists sense perception, memory, induction and logical reasoning.²³⁵ But none of these abilities can be attributed to the brain, according to Plantinga:

According to current science, electrons and quarks are simple, without parts. Presumably neither can think – neither can adopt propositional attitudes; neither can believe, doubt, hope, want, or fear. But then a proton composed of quarks won’t be able to think either, at least by way of physical relations between its component quarks, and the same will go for an atom composed of protons and electrons, a molecule composed of atoms, a cell composed of molecules, and an organ (e.g. a brain) composed of cells. If electrons and quarks cannot think, we won’t find anything composed of them that *can* think by way of the physical interaction of its parts.²³⁶

A brain cannot think, and that is an ontological problem for Plantinga, not a question of reliability. How does Plantinga think that we can think? By virtue of the soul, which is an immaterial substance.²³⁷ This is, as Plantinga puts it, not something that the soul does in virtue of the functioning of its parts (it does not have parts) it is a basic feature of the soul, just as it is a basic and intrinsic property of the electron to have a negative electrical charge.²³⁸

One of the reasons for naturalists to be materialists, according to Plantinga (and this is quite probably true) is that they are evolutionists, they believe that whatever faculties humans have, those have been brought about through an evolutionary process.²³⁹ And then the question arises:

How could an *immaterial self* or soul evolve in this way? What sort of genetic mutation would result in an immaterial soul? Could there be a section of DNA that codes not for the production of proteins of a certain sort, but for an immaterial self? That seems unlikely.²⁴⁰

²³⁴ Plantinga, Alvin “Materialism and Christian Belief” in van Inwagen, Peter; Zimmerman, Dean (ed.) *Persons: Human and Divine* (2007) p. 105-113

²³⁵ Plantinga, Alvin *Where the Conflict Really Lies: Science, Religion, and Naturalism* (2011) p. 311-312

²³⁶ Plantinga, Alvin “Materialism and Christian Belief” in van Inwagen, Peter; Zimmerman, Dean (ed.) *Persons: Human and Divine* (2007) p. 107-108

²³⁷ Plantinga, Alvin “Materialism and Christian Belief” in van Inwagen, Peter; Zimmerman, Dean (ed.) *Persons: Human and Divine* (2007) p. 116

²³⁸ Plantinga, Alvin “Materialism and Christian Belief” in van Inwagen, Peter; Zimmerman, Dean (ed.) *Persons: Human and Divine* (2007) p. 117

²³⁹ Plantinga, Alvin & Tooley, Michael *Knowledge of God* (2008) p. 33

²⁴⁰ Plantinga, Alvin & Tooley, Michael *Knowledge of God* (2008) p. 33

The problem with materialism is that it seems impossible that something material can be conscious, acquire propositional content and think. But the soul, which is supposed to be responsible for these capacities, is not, according to Plantinga, likely to be the product of genetic mutations. The problem is not that mutations are random, it is a more fundamental (a metaphysical) problem: souls do not arise by way of recombination of matter. Given this view of the mind-body problem a guided evolution is of no help. What you need to get cognitive faculties are immaterial souls, according to Plantinga. Then it seems as if this whole problem of “unguided evolution” melts away, in light of the more fundamental problem: that evolution cannot give rise to cognitive faculties (as opposed to indicators).

3.8.1 The proposed solution as inadequate for a what Plantinga calls “materialist”

But let us say we do think that material structures can be conscious, have beliefs with propositional content and can think, so that these abilities might have been brought about through an evolutionary process. For whom is the proposed solution – guided evolution - then a possible solution? Presumably a theist, who is what Plantinga calls a “materialist”. But a central premise for the argument is that an unguided evolution is not likely to give rise to reliable cognitive faculties, since beliefs cannot enter the causal chain leading to behavior, given materialism. That is why we need to complement the theory of evolution with the belief that God has directed this process so as to realize his goal of creating “knowers”.

But if God did so complement the evolutionary processes the resulting creatures would only be knowers, not actors. Why? If they were, that is, if their beliefs could be causally efficacious in virtue of their semantic content and their cognitive faculties had been brought about through a (guided) evolutionary process, then it seems as if even given materialism beliefs, *qua* beliefs, can, after all, enter the causal chain leading to behavior. And then a premise for the argument is false: given materialism beliefs *qua* beliefs can after all enter the causal chain leading to behavior.

If beliefs cannot enter the causal chain leading to behavior (given materialism), the implied solution would then mean that what God has created are observers, not actors. But this will hardly suffice to say that these beings are “made in the image of God” which is the main reason why Plantinga thinks that theists are warranted in believing that our cognitive faculties are reliable. But “*imago Dei*” implies also that we are actors.²⁴¹ So, neither for Plantinga, according to which the processes of evolution (guided or unguided) cannot give rise to thinking creatures with conscious beliefs, nor for the materialist theist, the implied solution is adequate.

But Plantinga’s argument is still important, because it shows that if we want to escape the conclusion that our cognitive faculties are unreliable, and we do not want to espouse God as an additional factor alongside those factors studied by evolutionary theory, we must allow beliefs to enter the causal chain leading to behavior, and this is a controversial conclusion. The reliability of our cognitive faculties in the defined sense, is a prerequisite for them being trustworthy. But to make sense

²⁴¹ As Plantinga points out in *Where the Conflict Really Lies* (2011) p. 269

of that reliability, to “locate” it within the naturalist grand story presupposes that this reliability has “contributed to (...) the struggle for reproductive advantage, more specifically, because they contributed to the tasks of feeding, fighting, fleeing and reproducing”²⁴² but in order to do so the content of our beliefs must make a difference to our behavior, in which case mental causation *qua* mental must be real.

But if all causation that takes place takes place in virtue of the traditional physical properties: momentum, charge, mass or energy, then, of course, propositional content is not part of the causal story of anything in the world. Unless the materialist agrees that propositional content *is* a part of the causal story, *qua* propositional content, the materialist will be faced with the problem of semantic or propositional epiphenomenalism, which Plantinga shows leads to a self-defeating position. So the conclusion is that the naturalist must have a richer view of nature than most physicalists have, a view in which mental causation is real.

Chapter 4. Conclusions and questions for further research

4.1 Summary and conclusions

The focus of this essay has been the presentation and the analysis of two influential arguments, Jaegwon Kim’s so-called supervenience argument, which aims to show that non-reductive materialism cannot give room for mental causation, and Alvin Plantinga’s evolutionary argument against naturalism, which aims to show that given a naturalistic evolution of our cognitive faculties, we have no reason to believe that our cognitive faculties are reliable, but that we, if we believe God has guided the evolutionary process are rationally justified in so believing.

I believe that both Kim and Plantinga are right in their analysis of their respective problems. In Kim’s case: that non-reductive physicalism is not compatible with mental causation, unless one abandons the principle of the causal closure of the physical domain. However I believe that what Kim does not see is that in order to get mental causation *qua* mental, it is not enough to reduce mentality to physical properties, one must do so in a way that preserves these mental properties as mental. Since ontological reduction is identification and identities are symmetrical, it implies a radical expansion of one’s understanding of the physical to include what is traditionally called mental properties and mental causation. Kim’s own solution, functional reduction is not adequate, as it does not give room for mental causation as he (and I) understand it. Kim effectively eliminates the mental properties that are supposed to be causally efficacious in mental causation.

Kim’s functional reduction involves two steps: first a reduction as translation in which the conscious (and therefore defining) character of mentality is defined away – it is eliminated – being

²⁴² Moreland, James Porter “The Argument from Consciousness” in *Blackwell Companion to Natural Theology* (2012) the expression “at least not hindered” has been left out in this quotation, because the reliability of our cognitive faculties would then be a brute, unexplained fact.

defined as a certain sequence of behavior or a function. The second step is to find the physical realizers for this behavior: neurophysiological properties that are the physical realizers of the behavior in question. In this reduction Kim is not able to save mental causation *qua* conscious, and that was the original motivation for why it is so important, according to Kim, to physically reduce mental phenomena:

“First and foremost, the possibility of human agency, and hence our moral practice, evidently requires that our mental states have causal effects in the physical world.”²⁴³

And:

“I have argued that mental causation requires reduction, and that anyone who believes in mental causation must be prepared to endorse mind-body reduction.”²⁴⁴

This discussion also made it clear that there is an important difference between giving a causal account of consciousness in non-mental terms and ontologically reducing consciousness to something non-mental. Ontological reduction means identifying - in this case identifying consciousness with something non-mental and therefore eliminating the mental. The second part of Kim’s functional reduction consists in providing a causal explanation in terms of neurophysiology of that which the first part has defined. If Kim had not defined away consciousness in the first part, the second part would not have led to an ontological reduction of consciousness to something physical, only to a causal reduction. And that is not enough for Kim, because it would have put him in the same place where he began: with mental properties being caused by or supervening on physical properties.

Kim’s critique of property dualism is that it makes the mental *qua* conscious causally inefficacious unless you give up the causal closure of the physical domain. The absurdity is that in his eagerness to save mental causation Kim is happy to throw mentality *qua* consciousness overboard. Kim acknowledges that not all mental phenomena are possible to define in non-mental terms and that in so far they exist – and Kim agrees that some do – physicalism does not survive and these irreducible mental phenomena are impotent.²⁴⁵

It seems as if there is no middle ground between elimination and panpsychism, if you want to ontologically reduce mentality to the physical domain. In the last section of *The Myth of Nonreductive Materialism* Kim argues that non-reductive materialism is not a stable position, that it will tend towards eliminativism via epiphenomenalism (if one takes causality as a guide to ontology, one can eliminate the mental which is impotent given non-reductive materialism, if Kim is right) or outright dualism, if you accept that the mental can exert causal power.²⁴⁶ I believe that my discussion shows that reductive materialism is not a stable position: it will tend towards panpsychism or eliminativism.

Plantinga’s argument also rightly points out a problem and that is that given that our cognitive faculties have evolved through the processes that evolutionary biology tells us about, mental causation

²⁴³ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 9

²⁴⁴ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 161

²⁴⁵ Kim, Jaegwon *Physicalism, or Something Near Enough* (2005) p. 162

²⁴⁶ Kim, Jaegwon *The Myth of Nonreductive Materialism* (1989) p. 47

qua mental must take place in order for us to have a chance of making the claim that our cognitive faculties are reliable, that is, tend to produce true beliefs when given true information. But in order to get mental causation we must, as Kim's argument shows, either abandon the causal closure of the physical domain - if the mental is not reducible to the physical - or expand our understanding of the physical so as to make room for mental causation *qua* mental. But any of these two steps would mean that we have essentially abandoned physicalism – either by abandoning the principle of the causal closure of the physical domain, or by expanding our understanding of what the physical domain contains, namely also mental properties exercising causal powers *qua* mental.

Therefore, if we combine Plantinga's and Kim's analyses we get an evolutionary argument against physicalism. The argument has the following structure:

Premise 1: Physicalism means that everything that exists is physical (reductive physicalism) or that everything that exists either is physical or supervenes on something physical (non-reductive physicalism).

Premise 2: Physicalism implies a limitation to one's ontology: our understanding of matter will not be so radically changed so as to include mental *qua* mental properties among its basic properties.

Premise 3: Panpsychism, according to which matter has among its basic properties mental, conscious properties, is therefore given premise 2 not a form of physicalism.

Premise 4: Both a reductive and non-reductive physicalist must accept the causal closure of the physical domain (that is: that every physical event that has a cause at *t*, has a physical cause at *t*).

Premise 5: Jaegwon Kim shows that non-reductive physicalism excludes mental (*qua* mental) causality) if one accepts the causal closure of the physical domain.

Premise 6: Plantinga's *Evolutionary Argument against Naturalism* shows that there can only be selection pressures towards reliable cognitive faculties if mental causation is real.

Premise 7: To argue against the reliability of our cognitive faculties presupposes the reliability of our cognitive faculties and is therefore implicitly self-defeating.

Conclusion:

If we believe that our cognitive faculties have evolved through the processes that evolutionary biology talks about, and we believe that they are reliable, we must either:

- A. If we are non-reductive physicalists: abandon the principle of the causal closure of the physical domain so that the mental can be causally efficacious *qua* mental.

OR:

- B. If we are reductive physicalists: expand our understanding of the physical domain so that it can include mental, conscious properties and their causal efficacy as part of the physical domain, that is, we must adopt a form of panpsychism.

Either of these two steps implies an abandonment of physicalism. Both of these versions are radical transformations of the most popular worldview among western philosophers, physicalism. Both of these versions are compatible with a theological anthropology that affirms that humans are knowers and capable of rational action, that these traits have evolved through the processes that evolutionary biology describes.

4.2 Questions for further research

The focus of this essay has been on the ontological implications of mental causation, and a further question is of course how one shall think of the “mechanics” of mental causation. The purpose of this essay has been to argue that mental causation happens and that we cannot deny it only because we do not yet have a good *explanans* of how mental causation works. The question of how mental causation takes place is of course both very interesting and hard and one which I have not addressed in this essay. In this essay the focus has been on asking if it exists, and a first step in the quest of understanding mental causation might be to acknowledge its existence.

Another area of reflection is the relationship between mental causation and free will. Keith T. Maslin points out the “holistic character” of intentional states, that beliefs are always related, in virtue of their content, to other beliefs, in a way that sensations are not. We may have an ache in our feet without that implying anything about sensations elsewhere in the body. But beliefs stand in logical relationships to other beliefs. My belief that there is an election on Sunday implies a lot of other beliefs.

Our self-consciousness makes it possible for us to evaluate these beliefs and see to what extent they fit together, and what is logically implied by one belief and how two beliefs fit together.²⁴⁷ This is of course what we do in any rational argument and this essay has been an attempt to evaluate the logical implications of two beliefs, namely that our cognitive faculties are reliable and that they have evolved through the processes of evolution.

In order to engage in such evaluation we presuppose that we have a capacity for rational evaluation and that capacity is in itself a part of what it means to have reliable cognitive faculties. But in order to be able to rationally evaluate our beliefs it seems as if our beliefs must be possible to form and withhold on logical and rational rather than merely physically causal grounds. This implies that what we think is not fully determined by physical causation. The further question then arises: what is the relationship between this freedom from physical determination and libertarian free will?

Mental causation is not yet free will, but beings with self-consciousness can morally evaluate their own desires, just as they can logically evaluate their own beliefs, as Harry Frankfurt points out in his *Freedom of the Will and the Concept of a Person*.²⁴⁸ Frankfurt talks about first- and second-order desires. We may desire to engage in revenge, but we may have a second-order desire that we did not have such a desire and therefore try to withhold our anger and also over time emotionally process our loss and try to cultivate another attitude. We can critically reflect, not only on our beliefs but also on our desires, not only on what we believe but also on what we want. Maybe free will is what comes through self-consciousness and the ability to reflect upon one’s beliefs and desires, and mental causation is what comes through consciousness.

²⁴⁷ Maslin, Keith T *An Introduction to the Philosophy of Mind* (2007) p. 25-26

²⁴⁸ Frankfurt, Harry “Freedom of the Will and the Concept of a Person” *The Journal of Philosophy*, Vol. 68, No. 1

References

Books

- Barbour, Ian *Religion in an Age of Science* Volume 1, San Fransisco: Harper & Row (1990)
- Beilby, James *Naturalism defeated? Essays on Plantinga's Evolutionary Argument against Naturalism* New York: Cornell University Press (2002)
- Bråkenhielm, Carl Reinhold *Filosofiska livsåskådningar* Uppsala: Uppsala Universitet (2008)
- Clayton, Philip *Mind & Emergence: From Quantum to Consciousness* Oxford: Oxford University Press (2004)
- Engdahl, Oskar; Larsson, Bengt *Sociologiska perspektiv: grundläggande begrepp och teorier* Lund: Studentlitteratur (2006)
- Gärdenfors, Peter *Hur Homo blev Sapiens: Om tänkandets evolution* Nora: Nya Doxa (2000)
- Holt, Nigel *Psychology: The Science of Mind and Behavior* Berkshire: McGraw Hill Higher Education (2012)
- Kim, Jaegwon *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation* Cambridge, Massachusetts: The MIT Press (1998)
- Kim, Jaegwon *Physicalism, or Something Near Enough* Princeton: Princeton University Press (2005)
- Kim, Jaegwon *Philosophy of Mind* Boulder, Colorado: Westview Press (2011)
- Maslin, Keith T. *An Introduction to the Philosophy of Mind* Cambridge: Polity Press (2007)
- Plantinga, Alvin; Tooley, Michael *Knowledge of God* Oxford: Blackwell publishing (2008)
- Plantinga, Alvin; Dennet, Daniel *Science and Religion: Are They Compatible?* New York: Oxford University Press (2010)
- Plantinga, Alvin *Warrant and Proper Function* New York: Oxford University Press (1993)
- Plantinga, Alvin *Where the Conflict Really Lies: Science, Religion, and Naturalism* New York: Oxford University Press (2011)
- Popper, Karl; Eccles, John *The Self and Its Brain* Berlin: Springer Verlag (1977)
- Runehov, Anne *Neuroscientific Explanations of Religious Experience: A Philosophical Evaluation* Uppsala: Uppsala Universitet (2004)
- Scruton, Roger *Kant – A Very Short Introduction* Oxford: Oxford University Press (2001)
- Searle, John R. *The Rediscovery of the Mind* Cambridge, Massachusetts: The MIT Press (1992)

Articles online

- Batterman, Robert "Intertheory Relations in Physics" in *Stanford Encyclopedia of Philosophy*
<http://plato.stanford.edu/entries/physics-interrelate/>
- Chalmers, David "Facing Up to The Problem of Consciousness" published in *Journal of Consciousness Studies* 2(3):200-19, (1995) available at: <http://consc.net/papers/facing.html>
- Cherry, Kendra "Formal Operational Stage of Cognitive Development"
<http://psychology.about.com/od/piagetstheory/p/formaloperation.htm>

Cole, David “The Chinese Room Argument” in *Stanford Encyclopedia of Philosophy*
<http://plato.stanford.edu/entries/chinese-room/>

Crane, Tim; Brewer Bill “The Mental Causation Debate” in *Proceedings of the Aristotelian Society Supplementary Volume LXIX* 1995 available at:
<http://sas-space.sas.ac.uk/218/1/The%20mental%20causation%20debate.pdf>

Crane, Tim; Mellor D.H “There is No Question of Physicalism” in *Mind* Vol. 99 (1990) at:
<http://www.jstor.org/stable/2254959>

Earl, Brian “What Does the evidence tell us about the biological value of consciousness?” in *Journal of Consciousness Studies*, 15, No. 7, 2008 http://www.imprint.co.uk/pdf/15_7-Earl.pdf

Frankfurt, Harry *Freedom of the Will and the Concept of a Person* in *The Journal of Philosophy*, Vol. 68, No. 1
<http://static.squarespace.com/static/5033029a84ae7fae2e6a0a98/t/5366a7a6e4b0c0f09ccda517/1399236518041/Frankfurt.pdf>

Hansson, Sven-Ove “Vetenskap och reduktionism” in *Folkvett nr 1/1991* available at
<http://www.vof.se/folkvett/ar-1991/nr-1/vetenskap-och-reduktionism/>

Kim, Jaegwon “Responses” *Philosophy and Phenomenological Research* Vol. LXV, No. 3, November 2002 Available at: <http://onlinelibrary.wiley.com/doi/10.1111/j.1933-1592.2002.tb00231.x/pdf>

Kim, Jaegwon “The Myth of Nonreductive Materialism” in *Proceedings and Adresses of the American Philosophical Association* available at: <http://www.jstor.org/stable/3130081>

Plantinga, Alvin “Advice to Christian Philosophers” *Faith & Philosophy* Vol. 1 Issue 3 (1994) available at: http://www.faithandphilosophy.com/article_advice.php

Plantinga, Alvin “Against Materialism” *Faith and Philosophy* Vol. 23, No 1 (2006) available at:
http://www.andrewmbailey.com/ap/Against_Materialism.pdf

Plantinga, Alvin “Content and Natural Selection” in *Philosophy and Phenomenological Research* Vol. 83 available at: http://www.andrewmbailey.com/ap/Content_Natural_Selection.pdf

Plantinga, Alvin “Evolution, Epiphenomenalism, Reductionism” in *Philosophy and Phenomenological Research* Vol. 68, No.3, May 2004 available at:
http://www.andrewmbailey.com/ap/Evolution_Epiphenomenalism_Reductionism.pdf

Plantinga, Alvin “Religion and Science” in *Stanford Encyclopedia of Philosophy*
<http://plato.stanford.edu/entries/religion-science/> (retrieved: 2014-06-09)

Ramsey, William “Eliminative Materialism” in *Stanford Encyclopedia of Philosophy*
<http://plato.stanford.edu/entries/materialism-eliminative/>

Sievert, Charles “Consciousness and Intentionality” in *Stanford Encyclopedia of Philosophy*
<http://plato.stanford.edu/entries/consciousness-intentionality/>

Smart, JJC “The Mind/Brain Identity Theory” in *Stanford Encyclopedia of Philosophy*
<http://plato.stanford.edu/entries/mind-identity>

“Solipsism” in *Nationalencyclopedia* <http://www.ne.se/solipsism>

Stoljar, Daniel “Physicalism” in *Stanford Encyclopedia of Philosophy*
<http://plato.stanford.edu/entries/physicalism/>

Strawson, Galen "Realistic Monism: Why Physicalism entails Panpsychism" in *Journal of Consciousness Studies*, Vol. 13, No. 10-11 (2006) Available at:
http://www.utoronto.ca/~seager/strawson_on_panpsychism.doc

Ney, Alyssa "Reductionism" in *Internet Encyclopedia of Philosophy* <http://www.iep.utm.edu/red-ism>

Robinson, William S "Evolution and Epiphenomenalism" in *Journal for Consciousness Studies*, 14 (11) (2007) available at: <http://www.public.iastate.edu/~wsrob/EvoEpi.pdf>

Walter, Sven "Epiphenomenalism" in *Internet Encyclopedia of Philosophy*
<http://www.iep.utm.edu/epipheno/>

Articles in books

Beckermann, Ansgar "Introduction – Reductive and Non-reductive Physicalism" in Beckermann, Ansgar; Flohr, Hans; Kim, Jaegwon (ed.) *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism* Berlin: Walter de Gruyter & Co (1992)

Block, Ned "Troubles with Functionalism" in *Readings in the Philosophy of Psychology* Vol. 1 Cambridge, Massachusetts: Harvard University Press (1980)

Moreland, James Porter "The Argument from Consciousness" in *The Blackwell Companion to Natural Theology* Oxford: Blackwell Publishing (2012)

"Orsak/verkan" in Lübecke, Poul (red) *Filosoflexikonet* Stockholm: Bokförlaget Forum (1988)

Plantinga, Alvin "Materialism and Christian Belief" in van Inwagen, Peter; Zimmerman, Dean (ed) *Persons: Human and Divine* Oxford: Oxford University Press (2007)

Swinburne, Richard "Själen behöver en hjärna för att fungera" in Hasker, William; Peterson, Michael; Reichenback, Bruce; Basinger, David (red.) *Religionsfilosofiska texter* Nora: Nya Doxa (1999)