This is a pre-print draft of a forthcoming chapter in Mihretu Guta (ed.), *Consciousness and the Ontology of Properties* (New York: Routledge). Please cite from the published version.

Neural Correlates of Consciousness & the Nature of the Mind

Matthew Owen

ABSTRACT: It is often thought that contemporary neuroscience provides strong evidence for physicalism that nullifies dualism. The principal data is neural correlates of consciousness (for brevity NCC). In this chapter I argue that NCC are neutral vis-à-vis physicalist and dualist views of the mind. First I clarify what NCC are and how neuroscientists identify them. Subsequently I discuss what NCC entail and highlight the need for philosophical argumentation in order to conclude that physicalism is true by appealing to NCC. Lastly, the simplicity argument for physicalism that appeals to NCC is presented, analyzed, and found wanting.

As one surveys the history of philosophy it is easy to get the impression that both materialism and dualism have been viable contenders.[1] From Antiquity to the Enlightenment, versions of dualism were considered rational options. Yet, modern philosophy has been dominated by various versions of materialism and physicalism (cf. Göcke, 2012, p. 1). In the early nineties, John Searle (1992, p. xiii) made the following assessment regarding contemporary philosophy of mind: "mainstream orthodoxy consists of various versions of 'materialism.'"

Today the nonphysical mind of dualism is commonly considered a relic of our pre-neuroscientific past. Many philosophers and neuroscientists agree with Nancy Murphy's (1998, p. 13) claim that neuroscience provides "dramatic evidence for physicalism." In *The Astonishing Hypothesis: The Scientific Search for the Soul*, molecular biologist and neuroscientist Francis Crick (1995, p. 3) proposes that human persons are "…in fact no more than the behavior of a vast assembly of nerve cells and

their associated molecules."[2] Even theologian Michael Horton (2011, p. 376) informs the faithful in *The Christian Faith: A Systematic Theology for Pilgrims on the Way* that neuroscience has firmly established "the mind is matter (i.e. the brain)."[3]

Granted, not everyone shares such sentiments. There are numerous dissenters. Nonetheless, most would still agree with a concession found in the *Stanford Encyclopedia of Philosophy*: "the 'neuroscientific milieu' of the past four decades has made it harder for philosophers to adopt dualism" (Bickle et al., 2010). True as that may be, we must ask: What is the evidence modern neuroscience provides for physicalism and against dualism? Surely it is not merely the fact that acts of the mind such as thought are associated with the brain. For this is hardly new information provided by modern neuroscience. The Greek physician Hippocrates (460-375 B.C.E.) knew this much (see 1886, p. 344), as did medieval philosopher and theologian, Thomas Aquinas (see ST 1a 78.4c). So what original evidence has modern neuroscience provided that disproves dualism?

The foremost answer is: neural correlates of consciousness (for brevity NCC). Accurate or not, the most common example of an NCC is C-fiber activation in one's brain that takes place when they are in a mental state of pain.[4] So when I am in a conscious state of pain there is a corresponding neural state in my brain, i.e. C-fiber activation. So the C-fiber activation is the neural correlate of the conscious state of pain, according to this example.

On the basis of such correlations, it is argued that physicalism is true and dualism is false. My aim is to explicate and analyze this line of reasoning. In the first section of this chapter, I will clarify what NCC are. Then the methods used to identify NCC will be

introduced in the following section, where I will briefly summarize some example studies. The third section will focus on what NCC imply. The fourth explains and critically analyzes the simplicity argument for physicalism based on NCC.

It should be noted that I will not concern myself with the question of whether there are NCC (cf. Noë and Thompson, 2004). I assume there are. My concern is what they imply, or what best explains them.

1 What is an NCC?

Simply put, a neural correlate of consciousness is a neural state or process that is correlated with consciousness. The idea is that when a subject is in a particular conscious state there is a corresponding state of their brain that is correlated with their conscious state. That is a basic description of an NCC.

However, in his influential work 'What Is a Neural Correlate of Consciousness?' David Chalmers (2000) points out that there are various conceptions of neural correlates within the NCC literature. Thus he tries to offer conceptual clarity by giving a theoretically neutral, reasonable definition that reflects common usage (2000, pp. 31, 38). As a starting point, Chalmers (2000, p. 17-18) presents and considers a definition of an NCC derived from the conference program of the Association for the Scientific Study of Consciousness.

> A neural system N is an NCC if the state of N correlates directly with states of consciousness.

Chalmers (2000, p. 18) then asks two clarifying questions: "First, what are the relevant 'states of consciousness'? Second, what does it mean for a neural state to 'correlate directly' with states of consciousness?"

Regarding the first question, Chalmers (2000, pp. 18-23) surveys several classes of phenomenal consciousness sometimes considered in the NCC literature. The first class is being conscious. A relevant NCC would be a neural state that correlates with a subject being conscious versus not being conscious. The second is a background state of consciousness – such as being awake, asleep, under hypnosis, in a state of flow, or the like. A corresponding neural correlate would be a neural state that directly correlates with one being under hypnosis.

The third class Chalmers covers is contents of consciousness (2000, p. 19). Suppose that after a day of teaching I come home from campus with roses. My wife will delightfully rush over and smell them. When she does, her experience of the smell of the roses is a specific content of her consciousness. A neural state that correlates with that particular experience of the smell of the roses would be a relevant NCC. Contents of consciousness are more fine-grained than the previous classes. The final class Chalmers considers is arbitrary phenomenal properties (2000, p. 22). Specific states of any of the above classes can be members of this class, which might be useful if one tries to give a general definition of an NCC.

When it comes to the second question, the complexity is even more apparent. The original question is: "What does it mean for a neural state to 'correlate directly' with states of consciousness?" Yet this question prompts Chalmers (2000, p. 24-28) to ask two more fundamental questions. First, must the neural state be necessary, sufficient, or necessary and sufficient for the conscious state it is correlated with? Second, must the correlation hold across all cases or only across specific types of cases (i.e. cases with ordinary brain function in an ordinary environment, cases with a normal brain but

unusual inputs, cases with varying stimulation, or cases with abnormal brain function due to lesions)?

Adequately answering the above questions is beyond the scope of this section. And my aim is not to formulate an original definition of an NCC. My ultimate aim is to show that a substance dualist position, neo-Thomistic hylomorphism, provides a good account of NCC. Such an account will be most effective if it assumes a reasonable definition that accords with general usage and is theoretically neutral. Since Chalmers's (2000, pp. 31, 38) definition is intended to meet such objectives, I will adopt his definition.

> An NCC is a minimal neural system N such that there is a mapping from states of N to states of consciousness, where a given state of N is sufficient, under conditions C, for the corresponding state of consciousness.[5]

While this definition is more precise, there are several parts that need explaining. First, the phrase '*minimal* neural system' needs clarification. Here Chalmers (2000, p. 24) is trying to avoid irrelevant neural processes being included (cf. Koch et al., 2016, p. 308). To clarify this point, let's consider a 'minimal engine system' that is a correlate of a car starting. The car starts when the ignition switch turns. So the turning of the ignition switch is a correlate of the car starting. However, there are other conditions true of the car when it starts. The gas tank will contain gas. The fuel line will be clear. The spark plugs will be clean. The crankshaft will be in place. The list goes on and on. Yet, if I were to explain to a new driver which of the above is a relevant correlate pertaining to their car starting, I need not explain the entire system of a properly functioning car engine. I only need to tell the new driver about the 'minimal engine system' that is a correlate of the car starting – i.e. when the ignition switch turns, the car starts.

Likewise, when it comes to an NCC we are not concerned with everything taking place in the entire nervous system, or even the brain in particular, when one is in a particular conscious state. Rather, we are concerned with the minimal neural state(s) or process(s) that correspond to that conscious state. To refer again to the common example regarding the mental state of pain and C-fiber activation, suppose my dog Anselm accidently bites my hand while playing. Is every firing synapse between the bite on my hand and my brain a neural correlate of my pain state? No, rather the neural correlate is the minimal neural system that, under certain conditions, is sufficient for me to feel the corresponding state of pain. According to our example, that would be the corresponding C-fiber activation, which is the minimal neural system since there is no more fundamental system that suffices for the corresponding state of pain (cf. Chalmers, 2000, p. 25). This is just one hypothetical example. Each minimal neural system will vary depending on the conscious state it correlates with.

The second part of Chalmers's definition that needs explanation is the qualifier 'under conditions C'. To return to our car analogy, while the ignition switch turning is the correlate of the car starting, there are further conditions true of the engine when the car starts. As mentioned above, the gas tank will have gas, the fuel line will be clear, the spark plugs clean, and so on. Such are conditions of a normally functioning car engine. And according to Chalmers (2000, p. 31), the conditions typically relevant to NCC include normal brain functioning that permits some atypical inputs and brain stimulation but not changes to brain structure (e.g. lesions).[6]

Lastly, let's clarify the phrase 'there is a mapping from states of N to states of consciousness.' First of all, this is not meant to suggest that there is only an NCC if we

have already identified it and mapped it. Rather, there is a mapping from the neural state N^1 to the conscious state C^1 if the former corresponds with the latter so that the correspondence could be mapped once identified. Secondly, the idea of mapping between corresponding neural states and conscious states pertains to subjects across a species, not just an individual subject. However, this mapping across a species is not necessarily a correspondence of identical neural states in every subject in a particular conscious state. The search for NCC is a search for biological regularities, and not necessary identical correspondence relations. Biological regularities of all kinds permit variations.

For elucidation, let's return to the familiar example of C-fibers and pain. Regarding the human species, there is a mapping from C-fiber activation to the conscious state of pain if it is true that when humans experience pain their conscious experience corresponds with C-fiber activation in their central nervous system. This does not, however, rule out variation. After all, pain can be one aspect of someone's overall conscious experience that includes additional mental states, which might result in neuronal variations. For example, some endurance athletes mentally train themselves to have an unusually high tolerance for pain through self-talk that affects their overall conscious experience when they are in pain. Thus while their experience of pain will be similar to the experiences of other subjects in pain, there will be some mental variation that could result in variation with respect to the neural processes.

Such variation is also relevant to a methodological challenge regarding controlled experiments. That is, it is very difficult if not impossible to produce identical overall conscious experiences in subjects being tested. If a neuroscientist showed me an image of red roses I could report to her my conscious perception of the image, just as my wife

could if shown the same image. Yet, my overall conscious experience will likely vary from hers even though we both consciously perceive the same image. After all, I do not care much for red roses whereas she absolutely loves red roses and gets very excited about them. And even if the neuroscientist could get us to have the exact same overall conscious experience, our brains are not exactly similar. In fact, no two individuals have brains that are exactly alike, not even identical twins, or even clones. In light of such variations, we should not expect the search for NCC to reveal correlations that are exactly the same across a species, but rather similar correlations reflecting biological regularities that permit variation.[7]

In sum, according to Chalmers's definition, an NCC is a minimal neural system that is sufficient under certain conditions for the corresponding state of consciousness, such that this correspondence can be mapped. Before concluding this section, it is worth noting that Chalmers's definition can be modified to apply specifically to specific types of phenomenal consciousness. For example, Chalmers (2000, p. 31) gives a modified definition particularly relevant to contents of consciousness.

> An NCC (for content) is a minimal neural representational system N such that representation of a content in N is sufficient, under conditions C, for representation of that content in consciousness.

While this modified definition applies to contents of consciousness, similar modifications could be made so that the definition applies specifically to other types of phenomenal consciousness. At times, such modifications might even be necessary.

Now that we are equipped with a definition of neural correlates of consciousness, let's consider how we identify NCC. The next section introduces standard methodology used to identify neural correlates of consciousness.

2 Identifying NCC

The aim of this section is to introduce methods used to identify neural correlates of consciousness. In meeting this objective I will refer to several example studies. The first study took place in the nineteenth century and provided a theoretical basis for techniques vital to the contemporary search for NCC. The second study, published in the year two thousand, pertains to neural correlates of binocular rivalry. It gives us an example of standard contemporary methodology used to identify NCC, which explicitly relies on subjective reports from study participants. The third study also pertains to neural correlates of binocular rivalry, but it implements the recently developed 'no-report paradigm.' This paradigm includes trials of the study with explicit reports from participants, as well as trials without explicit reports where physiological measures (e.g. pupil dilation) are used to infer what participants perceive (Koch et al., 2016, p. 308).

Fundamental to identifying neural correlates of consciousness is finding neural activity that consistently corresponds with certain conscious states. Imaging brain activity is central to this endeavor. Over a century ago Italian physiologist, Angelo Mosso (1846-1910), laid the conceptual basis for brain imaging techniques vital to the contemporary search for NCC (see Sandrone et al., 2014). A chief challenge to studying the brain is that it is enclosed in a hard protective casing – i.e. the skull. Mosso worked with a patient named Bertone who suffered extensive damage to the top of his skull, consequently much of it was missing. Where Bertone's skull was missing, Mosso placed a cap made out of a rubber-like substance, *gutta percha* (Glickstein, 2014, p. 343). This flexible cap made it possible to record brain pulsations of blood pressure correlated with mental activity such as emotional arousal and doing arithmetic (Glickstein, 2014, p. 343).

Mosso's study confirmed a straightforward hypothesis. That is, if the brain works harder there will be increased blood flow to the brain, so if the brain works harder when the mind works harder, there will be increased blood flow to the brain when the mind works harder. Put differently, (a) increased mental activity means (b) increased brain activity, which means (c) increased blood flow in the brain. So (a) increased mental activity correlates with (c) increased blood flow in the brain. Mosso confirmed this by measuring the increased blood flow in Bertone's brain that took place when Bertone's mental activity increased.

However, the method Mosso used to measure pulsations in Bertone's brain had a limitation. It was effective only if the patient had an abnormal skull breach (Sandrone et al., 2014, p. 622). Mosso's ingenious 'human circulation balance' was invented to overcome this limitation (Sandrone et al., 2014, p. 622). Mosso had his patients lay on a table that was essentially a balance intended to measure pulsations of blood flow that would tip the balance. Whether or not Mosso's human circulation balance was reliable, his work laid the conceptual foundation for noninvasive functional brain imaging techniques (Sandrone et al., 2014, p. 621-622). Noninvasive functional brain imaging is done while the brain is active and without being invasive to the brain by penetrating it in any significant way. Such brain imaging is vital to the search for NCC, since it allows us to see brain activity that corresponds with a conscious subject's mental activity.

Needless to say, noninvasive functional brain imaging technology has advanced significantly since Mosso's day. In the 1920s a German psychiatrist, Hans Berger, discovered it is possible to record electrical activity in the brain from the human scalp (Glickstein, 2014, p. 338). This type of recording is called an electroencephalogram

(EEG) (Glickstein, 2014, p. 338). Positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) have also since developed. A PET scan can reveal blood flow, glucose metabolism, oxygen metabolism, or concentrations of dopamine transporter indicative of brain activity (Johnson and Becker, 1999). An fMRI reveals increased brain activity by revealing increased blood flow. The blood carries oxygen on molecules of hemoglobin containing iron that changes the thermodynamic and magnetic properties of the brain area, which is detected by magnetic resonance imaging (see Bulte, 2011, 4:15-5:00).

Such technology is invaluable in the contemporary search for NCC, principally instigated in the late twentieth century by Francis Crick and Christof Koch (1990). To elucidate contemporary methodology, let's consider a conventional study done by Alex Polonsky and company (2000) to identify neural correlates of perception during binocular rivalry. In Polonsky's study fMRI was used to measure fluctuations of cortical activity correlated with alternating perceptions during binocular rivalry (2000, p. 1153). Binocular rivalry is a controllable perceptual illusion that takes place when both eyes are presented with different stimuli and the brain only allows one to be perceived (Mormann and Koch, 2007).

While two different rival stimuli were presented to each eye of study participants, the researchers measured fMRI signals in the early visual cortex (2000, p. 1153). One stimulus was a higher contrast green grating image. The other was a lower contrast red grating image. The subjects would report which stimulus they perceived by pressing one button when they perceived the green image and another button when they perceived the red image (2000, p. 1154). There was also a third button the subjects were instructed to

push when their perception of either the higher or lower contrasts were less than seventy-five percent homogeneous at particular times (2000, p. 1154). The researchers found that the fMRI signal in the primary visual cortex (V1) correlated with the perceptions the subjects reported. "V1 activity tended to increase when subjects reported seeing the higher contrast green grating, and the activity tended to decrease when they reported seeing the lower-contrast red grating" (2000, p. 1155).

To clarify the degree of these activity fluctuations in the fMRI signal, they did a separate series of fMRI scans to measure V1 activity as the stimuli physically alternated (2000, p. 1155). Basically, they compared V1 activity during binocular rivalry to V1 activity during stimulus alternation. Confirming earlier accounts (see Heeger et al., 2000), it was found that neuronal activity in V1 followed the alternations of the stimuli (2000, p. 1155). And they found that the fluctuations of V1 activity during binocular rivalry were forty-five to eighty-three percent as large as fluctuations in V1 activity induced during the stimulus alternations (2000, p. 1155). Furthermore, V1 activity fluctuations were about equal to those in visual areas nearby, that is: V2, V3, V4v and V3a (2000, p. 1155).

Polonsky and company (2000, p. 1157) concluded that their findings suggest neuronal activity critical for binocular rivalry is expressed as early as the primary visual cortex, V1, in the case of human vision. As they (2000, p. 1153) acknowledge, this runs contrary to the view that such neural activity occurs predominantly in later visual areas, which prior studies indicated (cf. Leopold and Logothetis, 1996; Logothetis and Schall, 1989; Sheinberg and Logothetis, 1997). Simply put, their methodology consisted of

controlling the stimuli presented to each eye of the subjects to induce binocular rivalry and imaging V1 neuronal activity using fMRI as the subjects reported their perceptions.

Such methodology critically depends upon the subjects' awareness of their perceptions, because the subjects report what they perceive to the researchers. The researchers then infer that the mental perception being reported correlates with the neural activity at that time, which is identified via brain imaging. This reliance on subjective reports is worrisome because researchers might confuse neural correlates of the subject's mental state during the report for neural correlates of the mental state being reported.[8] In other words, when a subject (*PERCEIVES*) perceives a green grating image without considering the fact that she perceives the image she is in a different mental state than when she (*REFLECTS*) perceives the image *and* considers her perception of it. In the former case she simply has a first order awareness of what is perceived (i.e. the green image). In the latter case she has a second order awareness of her first order awareness, so her mental state includes perceiving the green image and being aware of the fact that she perceives the green image.

Given that, a worry arises. That is, when the subject reports their first order awareness then they are in a different mental state that includes the second order awareness. And consequently the identified neural "correlate" may actually be correlated with the mental state that includes the second order awareness (i.e. *REFLECT*) when it is thought to be correlated with the first order mental state of simply perceiving the green image (i.e. *PERCEIVES*).

This worry motivates what is called the 'no-report paradigm' (cf. Tsuchiya et al., 2015). This methodological paradigm includes study trials with explicit reports from

participants, as well as trials without explicit reports where physiological measures are used to infer what participants perceive (Koch et al., 2016, p. 308). In 2014, Stefan Frässle and a team of researchers did a study on binocular rivalry with a pioneering application of a no-report paradigm (see Koch et al., 2016, p. 318, note 36). Again the subjects were presented with different stimuli to induce binocular rivalry. Their no-report paradigm used the ocular motor reflex, optokinetic nystagmus, along with pupil size as "objective measures" of which stimulus was dominant (Frässle et al., 2014, p. 1739). These reflex measures together with fMRI purportedly allowed them to assess the NCC of binocular rivalry without active reports from the subjects (Frässle et al., 2014, p. 1738). Frässle and company compared these measures to trials with active reports from the subjects, which allowed them to test the "applicability" of the objective measures (Frässle et al., 2014, p. 1743). When all was said and done, they concluded that active report and possibly introspection were partly responsible for neuronal activation patterns that are typically observed (Frässle et al., 2014, p. 1743-1745).

The no-reports paradigm is an innovative approach. Yet its accuracy depends on the reliability of the alleged objective measures, which may be difficult to establish. In any event, this concludes our brief overview of the methodology used to identify NCC. As an introduction to this methodology we have briefly considered three example studies intended to identify NCC. Given the discovery of neural correlates of consciousness, it is fitting to ask what they imply.

3 Implications of NCC

While the evidence for neural correlates of consciousness may be clear, it is not obvious what such correlations entail regarding the nature of the mind. At this point a common

fallacy – *post hoc, ergo propter hoc* – is tempting to commit. This fallacy is committed when one infers that ϕ caused φ simply because ϕ is correlated with φ. If one inferred that Barack Obama winning the democratic nomination in 2008 caused the stock market crash of 2008 simply because there is a correlation between the two events, they would commit this fallacy. Correlation does not entail causation. One needs further rationale to infer that a correlation is best explained by (or suggests) a causal relation.

Likewise, a correlation by itself does not entail dependency, identity, or that one correlate is reducible to the other. Given that ϕ is correlated with φ, we need more information to justifiably conclude that ϕ depends on φ, is identical to φ, or is reducible to φ. Suppose that all around the world whenever any philosopher heard a knock on their door they found a packaged philosophy book on their doorstep. On the basis of this correlation alone we couldn't justifiably infer that the packaged books caused the knocks, lest we commit the fallacy mentioned above. But likewise, we couldn't infer that the door knocks depend on the philosophy books. Nor could we infer that the knocks are identical to the books or in some way reducible to them. Further information would be required to justifiably make such inferences.

In some cases where we have identified an NCC we might have additional data that justifies further inferences. For example, suppose that whenever Fern was in a mental state of remembering her childhood in Kansas, a particular part of her brain lit up. And suppose further that the same area lights up in Kathryn's brain whenever she remembers her childhood in Washington State. Moreover, presume the data related to Fern and Kathryn confirms numerous studies with many human subjects. Given this, we could know that Fern and Kathryn's mental states of remembering their childhood correlate

with neural activity in a particular part of the human brain.

From this correlation alone we couldn't infer that Fern and Kathryn's mental state caused the neural activity, depended on it, was identical to it, or reducible to it. However, suppose that Fern and Kathryn lost the part of their brain with the neural correlates and directly after this they could never again remember their childhoods. Given this, the data set would then include more than just the correlations. With this additional data it would be justifiable to conclude that their mental state of remembering their childhood was not only correlated with the neural activity, but also depended on it.

If a physicalist assumed physics is fundamental before mapping the correlation between Fern and Kathryn's mental activity and the corresponding neural activity, she would likely conclude the correlation implies that the mental activity depends on the neural activity before gaining the additional data. And she would be justified in doing so to the degree that her assumption was well justified. But she would not be arriving at this conclusion merely on the basis of the correlation. Rather she would be justifiably arriving at the conclusion on the basis of the correlation *coupled with her preexperimental assumption*. Similarly, a dualist could justifiably conclude that Fern and Kathryn's immaterial minds stood in some type of causal relation with the neural correlates, if she justifiably assumed that their mental states are not reducible to physical states. However the dualist, like the physicalist, would be arriving at her conclusion on the basis of the NCC and her preexperimental assumption.

Pessimists might roll their eyes at this point and remind us that everyone has pre-experimental assumptions. That is true. Nevertheless, that does not mean we cannot learn anything from NCC. It does mean, however, that it is important to analyze our pre-

experimental assumptions and to be aware of the justificatory role they play. And according to Chalmers (1998, p. 227), "once we recognize the central role of preexperimental assumptions in the search for the NCC, we realize that there are limitations on just what we can expect this search to tell us." Given this, it is fitting that Thomas Metzinger (2000, p. 4) writes in his influential volume *Neural Correlates of Consciousness*:

> However, mapping does not mean reduction. Correlation does not mean explanation. Once strict, fine-grained correlations between brain states and conscious states have been established, a number of theoretical options are still open. Additional constraints therefore will eventually be needed. Important questions are What is the true nature of these psychophysical correlations? Are we justified in interpreting them as causal relations? What additional constraints would have to be introduced in order to speak of law-like correlations…? Is a fully reductive account, or even an eliminativist strategy, possible?

Though Metzinger is no dualist (cf. Metzinger, 2003), he goes on to acknowledge:

> Assume that we find a strict and systematic correlation between a certain brain property or type of neural event N and the subjectively experienced phenomenal property of "sogginess" S. This is entirely compatible with Cartesian dualism: The underlying relation could indeed be a causal one, namely causal interaction between events in two ontologically distinct domains.
>
> In short, correlations do not entail causation, dependency, identity, or reducibility.

Therefore further philosophical argumentation beyond the empirical data of NCC is required to arrive at a justified conclusion about the nature of the mind.

4 Against Dualism

So far I have argued that NCC do not entail a particular view of consciousness. As Jakob Hohwy (2007, p. 461) puts it: "…the notion of 'correlation' does not by itself commit one to any particular metaphysical view about the relation between (neural) matter and

consciousness." Given this, a philosophical argument is needed to show that NCC suggest dualism is false.

In *Sensations and Brain Process*, J.J.C Smart (1959) aims to provide such an argument against dualism. Smart's influential argument is based on the principle of simplicity, also known as Occam's razor. Simplicity says that when there are multiple theories that sufficiently explain a data set the theory that includes the least unnecessary entities is preferable.[9] Simplicity is one theoretical virtue that helps us discern which theory amongst competing theories provides the best explanation.

In this section I will present Smart's simplicity argument and then give a twofold critical analysis. After presenting the simplicity argument I will critique Smart's conclusion by summarizing problems that have convinced many philosophers that it is false. Then I will highlight a handful of weaknesses regarding the argument itself. My focus will be on Smart's simplicity argument, but it should be noted that he is not the only materialist who argues on the basis of simplicity. Christopher Hill (1991) has provided a more recent simplicity argument against dualism in the second chapter of *Sensations: A Defense of Type Materialism*.

For two reasons, I find it best to focus on Smart's presentation of the simplicity argument. One, Smart's presentation is widely influential. Two, Hill's argument and conclusion are much weaker, in my opinion. According to Hill (1991, pp. 28-29), there are three different ways the simplicity argument can be formulated corresponding to the different ways simplicity can be understood. Hill (1991, pp. 29-39) argues that two of the formulations fail. He then presents his own formulation, which relies on simplicity for

aesthetic appeal rather than epistemic justification (1991, pp. 39-40). Hill (1991, p. 40) concludes:

> It seems, then, that my claim for the simplicity argument must be modest. I must not maintain that it can be used to establish that type materialism is probable, nor that it can be used to convert all rational beings to type materialism. Rather, I can claim only that the argument makes a case that will be found persuasive by people whose aesthetic intuitions cause them to attach importance to ontological simplicity. It is, of course, my hope that the reader will find on reflection that he or she belongs to this group.

Hill's argument and conclusion seem too modest to seriously threaten dualism. Unlike Hill, Smart argues on the grounds of simplicity that it is indeed probable that materialism is true and dualism is false. If his argument is cogent, the implications for dualism are clearly consequential. Henceforth I will focus on Smart's simplicity argument.

4.1 Simplicity Argument

On the basis of simplicity Smart (1959) argues for the thesis that sensations are identical to neural processes. Due to advances in brain imaging technology, the amount of NCC data has grown significantly since Smart published *Sensations and Brain Processes* in 1959. Yet his argument and conclusion presuppose that neural processes are correlated with sensations.

Smart concludes that mental sensations are nothing more than their neural correlates, i.e. neural processes in the brain. So pain, for example, is allegedly C-fiber activation according to an 'is' of identity (see 1959, p. 145). The mental state of pain is nothing over and above the neural process of C-fiber activation. "They" are identical and thus have all and only the same properties.[10] This thesis is referred to as Smart's brain process theory or Smart's identity theory. Specifically, it is a type identity theory since it

says sensation types (e.g. pain or tasting chocolate) are identical to types of neural processes. By contrast a token identity theory says particular instances of a sensation (e.g. the pain I now feel after stubbing my toe) are identical to particular neural processes (cf. Smart, 2007).

Smart (1959, p. 142) acknowledges that consciousness appears irreducible. But "for various reasons," writes Smart (1959, p. 142), "I just cannot believe that this can be so." For one, it seems to him (1959, p. 142) "frankly unbelievable" that everything could be explicable in terms of physics and biology *except* for dualism's irreducible conscious states. Smart refers to such states as "nomological danglers" because they do not fit in a completely physical world. To make matters worse, these danglers need laws linking them to the brain processes they are correlated with (Smart, 1959, pp. 143, 156).

So dualism allegedly includes these copious irreducible conscious states – i.e. nomological danglers – and the laws "whereby the 'nomological danglers' dangle" (1959, p. 156). The alternative is the identity theory, which does not include such odd ontological baggage. According to the identity theory there are only brain processes that turn out to be sensations upon empirical investigation. Thus the identity theory is simpler and therefore enjoys the theoretical virtue of simplicity. In other words, simplicity favors the identity theory over dualism. So the identity theory is the best explanation of NCC data. We might formally reconstruct the argument like so.

(SIMPLICITY)	All else being equal, the simplest explanation of a data set is the best explanation.
(NCC-DATA)	Neuroscience has discovered NCC.
(ID-SIMPLER)	Relative to dualism, the identity theory is the simplest explanation of (NCC-DATA).

∴ (ID-BEST EX) Relative to dualism, the identity theory is the best explanation of (NCC-DATA).

The first premise (SIMPLICITY) is assumed by almost everyone. The second premise (NCC-DATA) is a perfectly safe assumption (see section 2.2). The fact that there are NCC is the data to be explained. The third premise (ID-SIMPLER) is a key step in arriving at the conclusion. Dualism supposedly includes nonphysical mental states (i.e. "nomological danglers") and the relevant laws relating them to neural events, whereas the identity theory does not.[11] So the identity theory is a simpler theory than dualism, says the simplicity argument's proponent. Hence the conclusion (ID-BEST EX) that the identity theory is the best explanation of NCC data, and therefore it is probably true.

Now that the simplicity argument has been explicated, we can analyze its merit. The following subsection will offer an analysis of the conclusion, i.e. the identity theory. A brief overview of the identity theory's primary problems will be given. If the identity theory has serious problems, its simplicity is counterbalanced by such considerations. If its problems are insurmountable and therefore we have reason to think the theory is false, its simplicity is moot. After commenting on the viability of the simplicity argument's conclusion, I will critically analyze the argument for that conclusion.

4.2 The Identity Theory's Viability

The simplicity of Smart's identity theory and its consistency with the explanatory completeness of physics makes it attractive to anyone with physicalist sympathies. But to echo Dickens (1859), the last half century or so has seen the best of times and the worst of times for the identity theory. On the one hand, various versions have been endorsed or at least seriously considered by some of the brightest philosophers. On the other hand,

grave problems have threatened its viability, leading many philosophers to begin looking elsewhere for a tenable theory. In this section, I will summarize two such problems that have been very influential. These issues have received much attention elsewhere, thus I will merely summarize them here without elaborating on them or defending relevant arguments.

Perhaps the most influential argument against the identity theory hinges on the multiple realizability of the mental (see Bickle, 2016). The problem of multiple realizability, famously raised by Hilary Putnam (1967), consists of the idea that it is possible for mental kinds to be realized by multiple physical kinds (cf. Bickle, 2016). Consider again the idea that pain is correlated with C-fiber activation. Given this, the identity theory entails that pain requires C-fiber activation. However, it seems possible that subjects without the same neurophysiology as humans, and wholly without C-fibers, could be in pain (cf. Bickle, 2016).[12] This, however, could not be possible if pain is identical to C-fiber activation.

A second problem the identity theory faces pertains to *qualia*, which are the experiences of what it is like to be in a conscious state. The experience of tasting Swiss chocolate feels different than the experience of tasting dirt. The difference is a difference of qualia. What it is like to taste Swiss chocolate is different than what it is like to taste dirt. The possibility of zombies has been used to clarify the problem presented by qualia for the identity theory. David Chalmers (1996, p. 94) describes a zombie as: "someone or something physically identical to me (or to any other conscious being), but lacking conscious experience altogether." While there are no actual zombies, many philosophers agree with Chalmers that zombies are possible. However, if the identity theory is true,

zombies are not possible. For if the identity theory were true, something physically identical to David Chalmers would necessarily have the same conscious experiences he has.

Multiple realizability and qualia have been widely influential in motivating many to discard the identity theory and even reconsider reductive physicalism altogether. Though there are additional problems with the identity theory worth consideration, let us move on and analyze the argument for the identity theory.[13]

4.3 Simplicity Argument Analysis

In this section a critical analysis of the simplicity argument will be given, but first two points of clarification are in order. For one, the principle of simplicity and the simplicity argument are not the same thing. The argument utilizes (or is based on) the principle. The principle of simplicity is merely a premise in the simplicity argument; hence it is not the argument. Moreover, successful and unsuccessful arguments can have one or more true premises.

This brings me to the second point: The epistemic principle of simplicity is indispensable. Not only is it theoretically well justified, its track record demonstrates it is pragmatically helpful. In this section objections to the simplicity argument will be given, but these objections are not aimed at the principle of simplicity. The argument can fail to be cogent, even though one of its premises – the principle of simplicity – is true and indispensable.

As for my critical analysis of the argument, I will highlight four concerns. First I will discuss the fact there are additional theoretical virtues to consider. My second concern pertains to the scope of data being explained. The third is how one's

metaphysical presuppositions (what one considers background knowledge) affect the apparent plausibility of competing theories. Lastly I will discuss the simplicity argument's misrepresentation of dualism.

Theoretical Virtues

One weakness of the simplicity argument is that it rests upon one, and only one, theoretical virtue. Simplicity is significant, but it is not the only theoretical virtue. To use Mario Bunge's (1961, p. 120) terminology from *The Weight of Simplicity in the Construction and Assaying of Scientific Theories*, simplicity "competes" with additional "desiderata." There are additional theoretical virtues to consider, such as explanatory power, explanatory scope, fertility, accuracy, internal coherence and consistency with widely accepted theories and background knowledge.[14]

If the identity theory is the simplest theory, that is a significant point in its favor. However, by itself simplicity is not conclusive. If simplicity itself were conclusive, then we should go with a view like solipsism and conclude that there is just one mind that exists that is imagining NCC data along with everything else. It is hard to think of a simpler theory. But given that there is more to consider than simplicity, we need not commit ourselves to solipsism simply because it is simple. Likewise, simplicity itself is not enough to establish that dualism is false and the identity theory is true.

Unless, of course, all else is equal and the identity theory is on par with (or better than) dualism on every other score. If that is the case, then simplicity can act as an epistemic "tiebreaker" tipping the balance in favor of the simplest view. In other words, the simplicity argument requires another premise affirming the condition mentioned in premise one, which reads: *All else being equal*, the simplest explanation of a data set is

the best explanation. Thus the needed premise is: (EQUAL) All else is equal, or in favor of the identity theory. So we can give a second, more accurate, formal reconstruction.

(SIMPLICITY)	All else being equal, the simplest explanation of a data set is the best explanation.
(EQUAL)	All else is equal, or in favor of the identity theory.
(NCC-DATA)	Neuroscience has discovered NCC.
(ID-SIMPLER)	Relative to dualism, the identity theory is the simplest explanation of (NCC-DATA).
∴ (ID-BEST-EX)	The identity theory is the best explanation of (NCC-DATA).

While a justifiable appeal to simplicity requires (EQUAL), some might think it needs no defense and that it is quite safe to assume. However, (EQUAL) is not obviously true, but questionable. As mentioned above, multiple realizability and qualia present two significant problems for the identity theory, but neither presents a problem for dualism. And contrary to what is commonly thought, there are considerable arguments offered by leading philosophers for dualism (see Loose et al., 2018; Lowe, 2009; Swinburne, 1986). Granted, dualism has its own problems. Thus elsewhere I have argued that a particular dualist view is capable of overcoming dualism's most significant problems (see Owen, 2018). My point here, however, is merely that (EQUAL) is questionable. It is also where much of the debate actually lies. And those inclined to either dualism or physicalism will often consider this premise in the light of differing metaphysical presuppositions. Such metaphysical presuppositions significantly influence the apparent plausibility of (EQUAL), and thus their role is well worth considering.

Metaphysical Presuppositions

In *Contemporary Aristotelian Metaphysics*, Tuomas E. Tahko points out how our metaphysical views influence how we approach empirical conclusions. Using the identity theory as an example, Tahko (2012, pp. 40-41) writes:

> Now, the question is: what sort of empirical information could verify this identity claim? We certainly have ample information about what happens in our brains, yet few physicalists would claim that this is by any means enough to settle the debate. In fact, I think that it is fair to say that *no* amount of purely empirical information could settle the debate by itself…The stalemate in contemporary philosophy of mind amounts to just this: the a priori delimitation of the different possibilities available in explaining consciousness has not been completed, at least not in sufficient detail to convince the majority of philosophers. Many philosophers are convinced that it is impossible to explain consciousness in terms of the physical, whereas others think that this is the *only* possible explanation.

What are the possible explanations of NCC and which explanations are impossible? That is a metaphysical question.

Physicalists and nonphysicalists typically disagree about what is and is not a possible explanation of NCC. Consider a physicalist who thinks (a) the nature of causation precludes nonphysical causes of physical effect and (b) our conscious states cause physical bodily effects. They will inevitably think that conscious states that cause certain neurons to fire resulting in muscle contractions and the like cannot possibility be nonphysical. On the flip side, consider a dualist who thinks (c) physical entities do not have intentionality (i.e. of-ness or about-ness) due to their nature and (d) our conscious states often have intentionality. They will naturally think a physicalist explanation of NCC is impossible since it would delete intentionality. In each case, metaphysical positions affect what each person thinks is a possible explanation of empirical data.

The metaphysical presuppositions that one approaches the issue of NCC with can pertain to something specific like the nature of causation or intentionality. However, there are broader metaphysical presuppositions that can play a role as well. For example,

Smart's materialism seems to rule out the possibility that consciousness could be nonphysical. He (1959, p. 142) writes: "That everything should be explicable in terms of physics (together of course with descriptions of the ways in which the parts are put together – roughly, biology is to physics as radio-engineering is to electro-magnetism) except the occurrence of sensations seems to me to be frankly unbelievable." Clearly, if the nature of reality is completely physical, then it is simply not possible for conscious states correlated with neural events to be nonphysical. Nonphysical conscious states just do not fit in a completely physical world. For this reason Smart (1959, p. 142) refers to irreducible conscious states as nomological danglers.

Given Smart's metaphysical view about the nature of reality, we could expect him to write: "I am just unable to believe in the nomological danglers themselves, or in the laws whereby they would dangle. If any philosophical arguments seemed to compel us to believe in such things, I would suspect a catch in the argument" (Smart, 1959, p. 143). Immediately following, he admits: "The above is largely a confession of faith…" These are metaphysical presuppositions he starts with. Given his starting points, his conclusion is to be expected.

But what if a non-materialist metaphysical view is presupposed? What if classical theism, is true? If so, there is a nonphysical omnipresent and transcendent God; moreover this nonphysical conscious God pre-existed everything physical and is explanatorily prior to everything physical. So presupposing classical theism, if anything is odd and out of place and thus dangles on reality, it is physical entities. Nonphysical conscious states fit quite nicely on such a metaphysical view. At the end of the day, if materialism is true

then physicalist theories of NCC max out the possibilities and if materialism is false then dualist views of NCC are possible.

Here is the main takeaway: Metaphysical presuppositions play a vital role in determining the possibility and plausibility of particular views of NCC. In other words, metaphysical presuppositions play a crucial role in justifying or undermining a key premise in the simplicity argument. That premise is (EQUAL), which says the identity theory is as plausible or more plausible than dualism. Dualists and physicalists might, and often do, consider (EQUAL) in the light of differing presuppositions that make this premise seem likely or unlikely. So this required premise cannot be assumed and must be defended, and that is where the real difficult work begins. At that point, an appeal to simplicity becomes far from simple.

Data Scope

Another limitation of the simplicity argument is that it appeals to a quite narrow scope of data. The simplicity argument accounts only for the *existence* of NCC. It does not reference any data about the *nature* of NCC. As leading NCC researcher, Koch (2016a) pointed out to me in an email:

> Note that the NCC themselves are neutral from the point of view of physicalism/materialism or one of the various shades of dualism. Under any reading, consciousness will have physical correlates. The question is what are those correlates, where are they and what can and will they tell us about how consciousness is generated in the first place. They also do not speak to whether or not consciousness can be analyzed using reductionism or other mereological assumptions.

The data revealing that there are NCC is well established and will become more so as the contemporary search for neural correlates continues (cf. Shulman, 2013). This is interesting data, but the fact that NCC exist does not tell us much about their nature. And

as Koch points out, the real question is: What are the correlates, i.e. what is their nature? The answer to that question will tell us much more than the fact that the correlates exist.

If an argument appeals to empirical data, you want the scope of the data to be as wide as possible. The simplicity argument makes such an appeal. A key premise is:

(NCC-DATA) Neuroscience has discovered NCC.

The problem is that this premise says nothing about the nature of NCC. Thus another weakness of the simplicity argument is that it does not appeal to any empirical data beyond the existence of NCC that tells us about the nature of such correlates. And the nature, rather than the existence, of NCC are more likely to provide information that has implications for the nature of the mental. In other words, the simplicity argument appeals to an empirical data set that has a narrow and limited scope.

Physicalists are naturally inclined to think a wider data scope that includes the nature of NCC will just further support physicalism anyway. After all, they think their view accurately describes reality as it pertains to the mind and a wider data scope will only tell us more about reality in that area. But for the exact same reason, nonphysicalists might reasonably expect a wider data scope to confirm their view.[15] Again one's presuppositions can play a role. Here they influence expectations of what a wider data set will support. In any event, it is to the simplicity argument's detriment that it appeals to such a narrow scope of neuroscientific data.

In addition to the simplicity argument appealing merely to the existence of NCC, there is another weakness. The simplicity argument only addresses neuroscientific data. Yet, we have more data about ourselves than neuroscientific data. We have what Roderick Chisholm (1976, pp. 16-18) called "philosophical data." Such data includes our

capacity to reason, to think about things, to make choices, our persistence through time and bodily changes, our moral awareness, and so on (cf. Chisholm, 1976). If a theory provides the simplest explanation of the fact that there are NCC, that is a point in its favor. However, that does not say much about how well it can account for philosophical data about human persons.

In sum, a considerable weakness of the simplicity argument is that it appeals to a narrow scope of data. It only appeals to neuroscientific data about NCC when there is also philosophical data about human persons, and the neuroscientific data it appeals to is limited to the existence of NCC and says nothing about their nature. To conclude that dualism is false and the identity theory is true based on such a narrow data set seems hasty, even if the identity theory is simpler.

Dualism Misrepresented

The version of dualism targeted by the proponent of the simplicity argument is often a minimal version of dualism, simply the idea that mental states are irreducible and nonphysical. And in light of NCC it is thought that laws relating irreducible nonphysical mental states to neural events must be postulated. This sort of dualism seems to be the target of Smart's (1959, pp. 142-143) presentation and Christopher Hill's (1991, p. 20) presentation of the simplicity argument. It is easy to understand the motivation for targeting a minimal version of dualism. The idea is that if you can undermine the minimal tenet(s) common to all versions of dualism, then you undermine all versions of dualism.

However, there are several problems with this approach. First, the minimal version of a position is not always the most defensible version. Sometimes a more robust version of a position is more defensible precisely because it is more complex, or more

nuanced, or has more resources available to deal with problems and provide explanations. Secondly, the minimal version of dualism that is the target of the simplicity argument misrepresents basic elements of the strongest versions of dualism. In my view, this would be David Oderberg's (2005) 'hylemorphic dualism,' J.P. Moreland's (2018) 'Thomistic-like dualism,' or my own version of neo-Thomistic hylomorphism.[16] I have argued elsewhere that neo-Thomistic hylomorphism provides a good explanation of NCC (see Owen, forthcoming). And if any versions of dualism actually entail NCC, as Jaegwon Kim (2005, p. 124) has claimed, this would most likely be true of Moreland's view.

Yet even Cartesian versions of dualism are far more defensible than the version of dualism the simplicity argument targets (see Swinburne, 1986; 2013). According to many hylomorphic and Cartesian dualists, the soul is a simple substance that can be in numerous mental states, or modes. Thus there is one substance existing in different modes according to such dualist views, which is significantly different than the idea that there is a collection of numerous individual mental properties. The former is arguably much simpler than the latter, which seems to be the target of the simplicity argument.

Thirdly, proponents of the simplicity argument seem to think that dualism is a neuroscience theory meant to account for neuroscientific data. But most dualist philosophers are not dualists because they think dualism is the best theory for explaining NCC or any other neuroscientific data (cf. Moreland, 2011, pp. 33-34). Typically, the rationale for dualism is strictly or principally philosophical and not empirical. For example, the nonphysical mind of substance dualism is thought to be what grounds human essence (cf. Oderberg, 2005), personal identity (cf. Lowe, 2001), subjective experience (cf. Zimmerman, 2011), agency (cf. Plantinga, 2012), cognition (cf.

Fumerton, 2013), moral awareness (cf. Swinburne, 1986), and/or the ethical value of a human person (cf. Moreland and Rae, 2000).

Since the nonphysical mind of substance dualism is not presented as a neuroscientific theory by dualists or argued for as if it were, it should not be evaluated principally on how good of a neuroscientific theory it is. That is not to say dualist views cannot provide a good account of NCC. After all, as mentioned above, I have argued that one dualist position certainly can (see Owen, forthcoming). What I am saying, however, is that dualist philosophers typically do not justify their position by appealing to dualism as an empirical theory that best explains empirical neuroscientific data, but rather as an explanation of non-empirical facts about ourselves.

5 Conclusion

The claim that neuroscience provides powerful evidence for physicalism and that undermines dualism is allegedly supported by the discovery of NCC. Yet as we have seen, such correlations do not entail any particular view of the mind. Moving from the fact of NCC to the conclusion that physicalism is true and dualism is false is not an empirical step. Rather it requires philosophical argumentation, which the simplicity argument aims to provide, but fails to do so satisfactorily. Thus contrary to popular opinion, it is not clear how neural correlates of consciousness justify physicalism and invalidate dualism.

References

Aquinas, T. (2002) **The Treatise on Human Nature: Summa Theologiae 1a 75-89**. Indianapolis, IN: Hackett Publishing Company, Inc.

Bayne, T. (2010) **Unity of Consciousness**. Oxford, UK: Oxford University Press.

Bickle, J. (2016) "Multiple Realizability". In Zalta, E.N. (Ed.) **The Stanford Encyclopedia of Philosophy**. Spring 2016 ed.

Bickle, J., Mandik, P. and Landreth, A. (2010) "The Philosophy of Neuroscience". In Zalta, E.N. (Ed.) **The Stanford Encyclopedia of Philosophy**. Summer 2012 Edition.

Bulte, D. (2011) "Magnetic Resonance Imaging Explained". Oxford Medical Illustration.

Bunge, M. (1961) The Weight of Simplicity in the Contemporary and Assaying of Scientific Theories. **Philosophy of Science**, 28: (2): 120-149.

Chalmers, D.J. (1996) **The Conscious Mind: In Search of a Fundamental Theory**. New York, NY: Oxford University Press.

Chalmers, D.J. (1998) "On the Search for the Neural Correlate of Consciousness". In Hameroff, S.R.;Kaszniak, A.W. & Scott, A.C. (Eds.) **Toward a Science of Consciousness II: The Second Tucson Discussions and Debates**. Cambridge, Massachusetts, MIT Press 219-229.

Chalmers, D.J. (2000) "What is a Neural Correlate of Consciousness?". In Metzinger, T. (Ed.) **Neural Correlates of Consciousness**. Cambridge, Massachussetts, MIT Press 17-40.

Chisholm, R.M. (1976) **Person and Object: A Metaphysical Study**. Chicago, Illinois: Open Court.

Cooper, J.W. (1989) **Body, Soul & Life Everlasting: Biblical Anthropology and the Monism-Dualism Debate**. Grand Rapids, MI: Eerdmans Publishing Co.

Cooper, J.W. (2009) The Current Body-Soul Debate: A Case for Dualistic Holism. **Southern Baptist Journal of Theology**, 13: (2): 32-50.

Crick, F. (1995) **The Astonishing Hypothesis: The Scientific Search for the Soul**. New York, NY: Touchstone.

Crick, F. and Koch, C. (1990) Toward a neurobiological theory of consciousness. **Seminars in the Neurosciences**, 2: 263-275.

Dickens, C. (1859) A Tale of Two Cities.

Feigl, H. (1967) **The Mental and the Physical: The Essay and Postscript**. University of Minnesota Press.

Frässle, S., Sommer, J., Jansen, A., et al. (2014) Binocular Rivalry: Frontal Activity Relates to Introspection and Action But Not to Perception. **The Journal of Neuroscience**, 34: (5): 1738-1747.

Fumerton, R. (2013) **Knowledge, Thought, and the Case for Dualism**. Cambridge, UK: Cambridge University Press.

Glickstein, M. (2014) **Neuroscience: A Historical Introduction**. Cambridge, Massachusetts: MIT Press.

Göcke, B.P. (ed.) (2012) **After Physicalism**, Notre Dame, Indiana: University of Notre Dame Press.

Goetz, S. and Taliaferro, C. (2011) **A Brief History of the Soul**. Oxford, UK: Wiley-Blackwell.

Hacker, P.H.S. (2007) **Human Nature: The Categorial Framework**. Oxford, UK: Blackwell Publishing.

Hasker, W. (2010) "Persons and the Unity of Consciousness". In Koons, R. & Beeler, G. (Eds.) **The Waning of Materialism**. New York, NY, Oxford University Press 175-190.

Heeger, D.J., Huk, A.C., Geisler, W.S., et al. (2000) Spikes versus BOLD: what does neuroimaging tells us about neuronal activity? **Nature Neuroscience**, 3: (7): 631-633.

Hill, C.S. (1991) **Sensations: A Defense of Type Materialism**. New York, NY: Cambridge University Press.

Hippocrates (1886) **The Genuine Works of Hippocrates**. New York, NY: William Wood and Company.

Hohwy, J. (2007) The Search for Neural Correlates of Consciousness. **Philosophy Compass**, 2: (3): 461-474.

Horton, M. (2011) **The Christian Faith: A Systematic Theology for Pilgrims On the Way**. Grand Rapids, MI.

Jackson, F. (1982) Epiphenomenal Qualia. **The Philosophical Quarterly**, 32: (127): 127-136.

Johnson, K.A. and Becker, J.A. (1999) **Neuroimaging Primer** [online]. http://www.med.harvard.edu/AANLIB/hms1.html [Accessed December 23 2016]

Kim, J. (2005) **Physicalism, Or Something Near Enough**. Princeton, New Jersey: Princeton University Press.

Kim, J. (2011) **Philosophy of Mind**. Third Boulder, CO: Westview Press.

Koch, C. (2006) "Religion and Science: Trying to Nail My Colors to the Mast". Unpublished.

Koch, C. (2016a) **Personal Communication**.

Koch, C. (2016b) "Studying Consciousness in the Mouse". **The 39th Annual Meeting of the Japan Neuroscience Society**. Yokohama, Japan.

Koch, C. (2017) **Personal Communication**.

Koch, C., Massimini, M., Boly, M., et al. (2016) Neural correlates of consciousness: progress and problems. **Nature Reviews Neuroscience***,* 17.

Kripke, S. (1981) **Naming and Necessity**. Malden, MA: Blackwell Publishing.

Leopold, D.A. and Logothetis, N.K. (1996) Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. **Nature***,* 379: (6565): 549-553.

Logothetis, N.K. and Schall, J.D. (1989) Neuronal Correlates of Subjective Visual Perception. **Science***,* 245: (4919): 761-763.

Loose, J.J., Menuge, A.J.L. and Moreland, J.P. (eds.) (2018) **The Blackwell Companion to Substance Dualism***,* Oxford, UK: Wiley Blackwell.

Lowe, E.J. (2001) "Identity, Composition, and the Simplicity of the Self". In Corcoran, K. (Ed.) **Soul, Body, and Survival: Essays on the Metaphysics of Human Persons**. Ithaca, New York, Cornell University Press 139-158.

Lowe, E.J. (2009) "Dualism". In McLaughlin, B.P.;Beckermann, A. & Walter, S. (Eds.) **The Oxford Handbook of Philosophy of Mind**. New York, NY, Oxford University Press 66-84.

Metzinger, T. (2000) "Introduction". In Metzinger, T. (Ed.) **Neural Correlates of Consciousness: Empirical and Conceptual Questions**. Cambridge, Massachusetts, The MIT Press.

Metzinger, T. (2003) **Being No One: The Self-Model Theory of Subjectivity**. MIT Press.

Moreland, J.P. (2011) Substance Dualism and the Argument from Self-Awareness. **Philosophia Christi***,* 13: (1): 21-34.

Moreland, J.P. (2018) "In Defense of a Thomistic-like Dualism". In Loose, J.J.;Menuge, A.J.L. & Moreland, J.P. (Eds.) **The Blackwell Companion to Substance Dualism**. Oxford, England, Wiley Blackwell.

Moreland, J.P. and Rae, S.B. (2000) **Body & Soul: Human Nature & the Crisis in Ethics**. Downer Groves, Illinois: InterVarsity Press.

Mormann, F. and Koch, C. (2007) "Neural Correlates of Consciousness". **Scholarpedia: the peer-reviewed open-access encyclopedia**.

Murphy, N. (1998) "Human Nature: Historical, Scientific, and Religious Issues". In Brown, W.S.;Murphy, N. & Malony, H.N. (Eds.) **Whatever Happened to the Soul? Scientific and Theological Portraits of Human Nature**. Minneapolis, MN, Fortress Press.

Noë, A. and Thompson, E. (2004) Are There Neural Correlates of Consciousness? **Journal of Consciousness Studies**, 11: (1).

Oderberg, D.S. (2005) "Hylemorphic Dualism". In Paul, E.F.;Fred D. Miller, J. & Paul, J. (Eds.) **Personal Identity**. New York, NY, Cambridge University Press 70-99.

Owen, M. (2016) "Neuroplasticity & consciousness of neural correlates". **NeuroFutures 2016: Circuit Structure & Dynamics**. Allen Institute for Brain Science, Seattle, WA.

Owen, M. (2018) **Neo-Thomistic Hylomorphism Applied to Mental Causation & Neural Correlates of Consciousness**. Doctoral dissertation, University of Birmingham.

Owen, M. (forthcoming) Aristotelian Causation & Neural Correlates of Consciousness.

Plantinga, A. (2012) A New Argument against Materialism. **Philosophia Christi**, 14: (1): 9-27.

Polgar, T.W. (2011) Are Sensations Still Brain Processes. **Philosophical Psychology**, 24: (1): 1-21.

Polonsky, A., Blake, R., Braun, J., et al. (2000) Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry. **Nature Neuroscience**, 3: (11): 1153-1159.

Putnam, H. (1967) "Psychological predicates". In Capitan, W.H. & Merrill, D.D. (Eds.) **Art, Mind, and Religion**. University of Pittsburgh Press 37-48.

Sandrone, S., Bacigaluppi, M., Galloni, M.R., et al. (2014) Weighing brain activity with the balance: Angelo Mosso's original manuscripts come to light. **Brain: A Journal of Neurology**, 137: 621-633.

Schwartz, J.M. (1999a) "First Steps toward a Theory of Mental Force: PET Imaging of Systematic Cerebral Changes after Psychological Treatment of Obsessive-Compulsive Disorder". In Hameroff, S.R.;Kaszniak, A.W. & Chalmers, D.J. (Eds.) **Toward a Science of Consciousness III: Third Tuscan Discussions and Debates**. Cambridge, Massachusetts, The MIT Press 111-122.

Schwartz, J.M. (1999b) A Role for Volition and Attention in the Generation of New Brain Circuitry: Toward a Neurobiology of Mental Force. **Journal of Consciousness Studies**, 6: (8-9): 115-142.

Schwartz, J.M. and Begley, S. (2002) **The Mind & The Brain: Neuroplasticity and the Power of Mental Force**. New York, NY: HarperCollins.

Science, A.I.f.B. **Allen Brain Atlas** [online]. http://brain-map.org/ Seattle, WA Allen Institute for Brain Science [Accessed September 13 2017].

Searle, J.R. (1992) **The Rediscovery of the Mind**. Cambridge, Massachusetts: MIT Press.

Sheinberg, D.L. and Logothetis, N.K. (1997) The role of temporal cortical areas in perceptual organization. **Proc. Natl. Acad. Sci. USA**, 94: 3408-3413.

Shulman, R.G. (2013) **Brain Imaging: What It Can (and Cannot) Tell Us About Consciousness**. New York, NY: Oxford University Press.

Smart, J.J.C. (1959) Sensations and Brain Processes. **The Philosophical Review**, 68: (2): 141-156.

Smart, J.J.C. (2007) "The Mind/Brain Identity Theory". In Zalta, E.N. (Ed.) **The Stanford Encyclopedia of Philosophy**. Winter 2014 Edition ed.

Swinburne, R. (1986) **The Evolution of the Soul**. Revised Edition.New York, NY: Oxford University Press.

Swinburne, R. (2013) **Mind, Brain, and Free Will**. Oxford, UK: Oxford University Press.

Tahko, T.E. (2012) **Contemporary Aristotelian Metaphysics**. Cambridge, UK: Cambridge University Press.

Tononi, G. and Koch, C. (2015) Consciousness: here, there and everywhere? **Philosophical Transactions of The Royal Society B**, 370.

Tsuchiya, N., Wilke, M., Frässle, S., et al. (2015) No-Report Paradigms: Extracting the True Neural Correlates of Consciousness. **TRENDS in Cognitive Sciences**, 19: (12): 757-770.

Zimmerman, D. (2011) "From Experience to Experiencer". In Baker, M.C. & Goetz, S. (Eds.) **The Soul Hypothesis**. New York, NY, Continuum 168-196.

Notes

I would like to thank Nikk Effingham and Jussi Suikkanen for reading and commenting on multiple early drafts of this chapter. I am also greatly indebted to Christof Koch for meeting with me regularly to discuss my research and providing very helpful feedback on a previous draft. Our conversations during our monthly meetings have been invaluable. Aryn Owen and Mihretu Guta also meticulously read an earlier draft and provided helpful constructive criticism on my writing. I am also grateful to them for numerous conversations related to the ideas presented here.

[1] For a brief survey, see Goetz and Taliaferro (2011).
[2] Cf. Mitchell Glickstein (2014, p. 1).
[3] Horton seems to be disavowing dualism and advocating materialism. The previous line says: "Philosophical defenses of materialism seem increasingly substantiated by science" (2011, p. 376). Following that line, he says neuroscience has proven the mind is matter. However, on the next page he advocates what he calls "dichotomy," according to which the soul is distinct from the body and persists after bodily death (2011, p. 377). This seems very dualistic. Moreover, on the same page he has a footnote where he commends John W. Cooper's position, which Cooper calls "dualistic holism" (see Cooper, 1989, Ch. 10 section IV). And elsewhere Cooper (2009, p. 46) commends "Thomist dualism."
[4] C-fiber activation correlated with pain is the standard example in the philosophical literature; therefore I will often use this example. However, most neuroscientists "would *not* consider these a true content-specific NCC…The pain-NCC is higher-order somato-sensory cortex…" (Koch, 2017).
[5] NCC researcher Christof Koch (2016, p. 307) gives a similar definition: "The NCC are defined as the minimum neuronal mechanisms jointly sufficient for any one specific conscious percept."
[6] Chalmers (2000, p. 32) points out that lesion studies are often used to make inferences about NCC, but he thinks such methodology is flawed. According to Chalmers, "the identity of an NCC is arguably always relative to specific brain architecture and normal brain functioning, and correlation across abnormal cases should not generally be expected." What was an NCC can cease to be such when a lesion alters brain structure, thus warranting caution when making inferences from lesion studies. However, as Koch pointed out to me on October 16, 2017, lesion studies can nevertheless provide important information, especially when findings from lesion studies are coupled with findings from artificial stimulation studies in a healthy brain that corroborate the lesion studies (see e.g. Koch et al., 2016, p. 308).
[7] I am indebted to Koch for this point and the foregoing sub points elucidating the overall idea. In our conversation on September 11, 2017, he pointed out the variations pertaining to NCC due to variations of overall conscious experience and differences in individual brains. Koch is the President and Chief Scientific Officer at the Allen Institute for Brain

Science, which has endeavored to map the human brain and mouse brain (see brain-map.org). The Allen Institute's efforts to map the mouse brain involve cloned mice.

[8] I am grateful to Koch for clarifying the motivation of the no-report paradigm during our conversation at the Allen Institute, March 15, 2017.

[9] Koch (2006, p. 11) sums it up well: "…this principle states that of equally good explanations for a phenomenon, the best one is the simplest explanation that accounts for all the facts."

[10] However that is not to say, according to Smart, that 'pain' means the same as 'C-fiber activation' (see 1959, pp. 144-145). Accordingly, one cannot infer that pain is identical to C-fiber activation from the meaning of terms; it is allegedly empirically discovered.

[11] Another interpretation of Smart says irreducible conscious states are "danglers" since they are epiphenomenal (cf. Feigl, 1967; Polgar, 2011; Smart, 2007). This makes Smart's argument hinge on the idea that irreducible mental states are necessarily epiphenomenal. Since many dualists disagree, Smart would need to show that such states are necessarily epiphenomenal. Some physicalists think causal closure entails such. It is worth noting that Jaegwon Kim (2011, Ch. 4) has argued that because nonphysical mental states would be epiphenomenal we should reduce them to neural correlates. Kim relies on the causal pairing problem to support the premise that nonphysical states would be epiphenomenal (see 2011, p. 113, endnote 15). For in depth discussions of causal closure and causal pairing, please see Owen (2018).

[12] At this point, twin earth scenarios become relevant (see Kripke, 1981).

[13] Additional arguments include the inverted spectrum argument (see Chalmers, 1996, p. 99) and the knowledge argument (Jackson, 1982, p. 130). P.M.S. Hacker (2007, p. 252) provides a short interesting list of issues. The unity of consciousness also poses a serious problem, in my opinion (cf. Hasker, 2010). And such unity is acknowledged by NCC researchers, Koch (2016b) and Giulio Tononi (2015, p. 6). For an in-depth defense of the unity of consciousness, see Tim Bayne (2010).

[14] While most of these theoretical virtues are commonplace, fertility is less known. If a theory leads to further research opportunities, it has the theoretical virtue of fertility.

[15] It has been argued that the nature of certain neural correlates relevant to cases of neuroplasticity imply irreducible conscious states (see Owen, 2016; Schwartz, 1999a; b; Schwartz and Begley, 2002).

[16] *Hylomorphism* is often spelled 'hylemorphism,' as Oderberg spells it.