



# Why NCC research is not theory-neutral

Michael Pauen<sup>a,b</sup>  (m@pauen.com)

## Abstract

One of the reasons why the Neural Correlates of Consciousness Program could appear attractive in the 1990s was that it seemed to disentangle theoretical and empirical problems. Theoretical disagreements could thus be sidestepped in order to focus on empirical research regarding the neural substrate of consciousness. One of the further consequences of this dissociation of empirical and theoretical questions was that fundamental questions regarding the Mind Body Problem or the “Hard Problem of Consciousness” could remain unresolved even if the search for the neural correlates had been successful.

Drawing on historical examples, a widely held consensus in the philosophy of science, and actual NCC research we argue that there is no such independence. Moreover, as the dependence between the theoretical and the empirical level is mutual, empirical progress will go hand in hand with theoretical development. Thus, contrary to what the original NCC program suggested, we conclude that NCC research may significantly take advantage from and contribute to theoretical progress in our explanation and understanding of consciousness. Eventually, this might even contribute to a solution of the Hard Problem of Consciousness.

## Keywords

Hard problem of consciousness · History of science · Neural correlates of consciousness · Pain

*This article is part of a special issue on “The Neural Correlates of Consciousness”, edited by Sascha Benjamin Fink and Ying-Tung Lin.*

## 1 Introduction

Maybe one of the most striking features of the reemerging debate about the problem of consciousness from the 1970s on were growing disagreements particularly among materialists. Until the late 1960s almost all materialists agreed that the central issue regarding consciousness was the metaphysical question whether or not mental states are physical states. But from the mid-seventies on, an increasing number of philosophers started to argue for an epistemic understanding of the

---

<sup>a</sup>Berlin School of Mind and Brain

<sup>b</sup>Humboldt-Universität zu Berlin, Research Training Group 2386 “Extrospection”



problem. According to authors like Thomas Nagel, Frank Jackson, Joseph Levine, or David Chalmers the really “hard” problem of consciousness was epistemic: Even if mental states *were*, in fact, physical states, it could still be asked how this fact can be *explained* and *understood*. Moreover, this *Hard Problem* might turn out to be unsolvable.

Almost at the same time, it became clear that any progress in consciousness research would require an interdisciplinary effort, involving particularly neuroscientists, psychologists, and philosophers. But what could be expected from such a combined effort, if there was no consensus as to what the problem was and whether it was solvable at all?

In this situation, an agreement on something like the lowest common denominator could appear extremely attractive, and this is exactly what the proponents of the NCC program had on offer. The basic idea was to set aside the controversies resulting from different theoretical commitments and to focus on “theoretically neutral” data instead, in order allow “researchers of widely different theoretical persuasions” (Chalmers, 2000, p. 37) to contribute. Empirical facts regarding the neural substrate of conscious experience were needed anyway, no matter what one’s own theoretical commitments were or whether one would believe that these facts were sufficient for a solution of the problem of consciousness.

Here we will focus on the assumption that the search for the NCCs is “theoretically neutral”, such that this search can be pursued independently of the diverging theoretical commitments consciousness researchers might have. Our first point will be that theory-neutrality actually is an essential claim of the NCC program. And it is not difficult to see why. Theory-neutrality was a response to a specific situation in the science of consciousness where theoretical disagreements seemed to threaten empirical progress. It could thus appear as a reasonable idea to set theoretical disagreements aside for the time being in order to establish the facts.

Apart from individual remarks particularly in Chalmers’s papers on the NCC program, this view also emerges from the idea about the development of consciousness science which is shared by Christof Koch and David Chalmers. According to this idea, consciousness science will develop in three step steps, with NCC research as the first step. While this first step was supposed to be theoretically neutral, the following two steps, by contrast, were described as theoretically loaded: Functional explanations that were supposed to come as a second step, and a full-blown theory of phenomenal consciousness as the third one somewhere in the distant future. Finally, this interpretation also explains why Chalmers and Koch talk about the neural *correlates* of consciousness: Our interpretation goes along with the textbook understanding of correlation as a regular co-occurrence of two phenomena that does not imply a causal or a mereological connection between them.

All this seems to indicate that the NCC program is clearly defined and well justified, so why is there a need for a discussion, in the first place? The reason why we think that the NCC program with its assumption of theory-neutrality deserves some attention is that it might seem to contradict a widely accepted assumption in the philosophy of science, namely that there is a dependence between empirical

data and scientific theories. Moreover, theory-neutrality is also at variance with the actual scientific practice both in neuroscience and in psychology.

Thus, our second point is, that this dependence between empirical studies and theoretical commitments holds for NCC research as well, and that it works both ways: While empirical research on consciousness is guided by theoretical presuppositions, new empirical developments may have consequences that lead to a revision of theoretical presuppositions as well. This has an important implication: As empirical findings affect our conceptual, methodological, and theoretical commitments, they may lead to a constant process of revision which makes it highly unlikely that we will be able to chase down the neural correlates without improving our theoretical understanding of consciousness. Our claims will get additional support as the history of science shows that the development of identity statements like the famous one on water/H<sub>2</sub>O did not only require empirical studies. Rather, they also involved a competition between opposing theories, in this case between the Phlogiston- and the Oxygen-Theory.

Taken together, all these considerations lead us to conclude that the NCC program which was thought to sidestep problems of explaining and understanding consciousness can make a significant contribution to a solution of these problems, which might even have repercussions on the *Hard Problem of Consciousness*.

We will start with a short sketch of the original NCC program, including a discussion of what *theory-neutrality* means in Part 2. Part 3 will provide some historical background regarding the identification of water and H<sub>2</sub>O which has often been taken as the gold standard for the mind/brain identity claim. The case will illustrate how theoretical presuppositions guide the search for evidence that eventually helps to establish an identity claim. In Part 4 we will take the debate between cognitive and non-cognitive theories of consciousness as an example to show that NCC research requires theoretical presuppositions for the design of experiments, the selection of research questions and working hypotheses, and for the interpretation of the data. Finally, Part 5 will demonstrate that the relation between empirical results and theoretical presuppositions also works the other way round: Empirical results can help to decide conflicts between opposing theoretical presuppositions, but they can also help to generate new presuppositions – in this case regarding the phenomenology and function of mental states like pain. We conclude that this mutual interaction between theoretical presuppositions and empirical data may even result in a significant contribution of NCC research to a better understanding of consciousness.

## 2 Neural correlates of consciousness and theory-neutrality

According to Chalmers (1998, 2000) and Koch (1990, 1997; 2002; 2004), NCCs are the minimal neural conditions jointly sufficient for a specific conscious experience



(Koch et al., 2016), e.g., for an experience of pain. The most promising candidates for NCCs themselves were originally thought to be activity states in specific brain regions, e.g. in the anterior cingulate cortex and the somatosensory cortex in the case of pain (Rainville et al., 1997), or activity in the ventral stream of the visual system as a correlate for visual experience (Chalmers, 1998; Milner & Goodale, 2008). This hasn't changed very much since then with the only exception that – with the advent of multivariate pattern analyses (Haxby, 2012; Haynes, 2009) – specific distribution patterns of neural activity rather than overall activity states have come into focus. Also, certain types of global brain activity, e.g. 40 Hertz oscillations (Crick & Koch, 1997) or so-called “NMDA synapse activity” (Flohr, 1992) were suggested as NCC candidates in the 1990s, but play a minor role today.

## 2.1 Theory-neutrality

As already indicated, theoretical disagreements seemed to severely hamper empirical research and interdisciplinary cooperation probably from the 1980s on. In this situation, the answer of the NCC program was simple and convincing: If we want to make any progress in our understanding of the physical substrate of consciousness, we need neuroscientific data anyway. So, let's focus on the neural correlates of consciousness and set the methodological and theoretical disagreements aside for the time being.

Regarding the development of consciousness science, Koch and Crick envisage a three-step process with the search for the NCCs as the first step. It is followed by the description of the causal link between neural activity and conscious experience as a second step, and the explanation of phenomenal experience, that is, the solution of the Hard Problem as the third and final step (Koch & Crick, 2002).

A similar plan has been brought up by David Chalmers. Like Koch and Crick, Chalmers envisages a three-step development in the science of consciousness, where the NCC program is the first step. It is “a project that is *relatively tractable, clearly defined, and theoretically neutral*, one whose goal seems to be visible somewhere in the middle distance” (Chalmers, 2000, p. 38, emphasis added). The second step is a functional explanation of the mechanisms underlying consciousness, while the final third step may be a solution of the Hard Problem (Chalmers, 1998). In order to understand the difference between the steps and the extension of the entire program it is worth noting that Chalmers expects that “a century or two” will pass before the functional explanations of the second step replace the NCC program.

In order to better understand what, exactly, Chalmers has in mind when he talks about theory-neutrality, two questions matter: First, what does it mean for a line of research to be theory-neutral? Second, what are the theories or theoretical commitments, with respect to which NCC research is supposed to be neutral?

Regarding the first question, theory-neutrality of a certain line of research can mean either that the relevant theoretical commitments are among the *presuppositions* or that they are among the *consequences* of that research. As the *consequences* of a given investigation are difficult to determine and largely beyond the author's control, it wouldn't make much sense to make neutrality regarding possible theoretical consequences an essential part of the NCC program. Another reason against this move is that it would not contribute to the obvious goal of the NCC program: to allow researchers with a broad spectrum of diverging theoretical commitments to contribute to consciousness research. Unlike the theoretical presuppositions of a study, the consequences are independent of the researcher's commitments. That is why we assume that what matters primarily are the theoretical *presuppositions* of NCC research, even if it will turn out below in Part 5 that there is a mutual relationship between consequences and presuppositions.

In fact, it sounds quite sensible to understand theory-neutrality in this way if one wants to establish a program that is open for 'researchers of widely different theoretical persuasions': Obviously, my theoretical persuasions matter for the presuppositions of my studies, that is, for the methodological commitments I accept, for the research questions I regard as worth investigating, and for my interpretation of the data. By contrast, my persuasions do not (or should not) matter for the data themselves and they certainly do not matter for the consequences that these results may have in the future.

## 2.2 Theoretical presuppositions

Let's look at the famous Ether Wind Experiments by Michelson and Morley (1887) to illustrate this view of theory-neutrality or theory-dependence, respectively. According to our view, Michelson and Morley's theoretical commitments regarding the Ether Wind Theory should become apparent in the theoretical presuppositions of their experiments. And this is indeed the case: The Ether Wind Theory shapes the research question (does the relative movement of the ether wind affect the speed of light?), the hypothesis (the ether wind should affect the speed of light) but not the result, which was negative and eventually led to the dismissal of the Ether Wind Theory. The Relativity Theory, by contrast, is not among Michelson and Morley's theoretical commitments, even if it was the most important consequence of their experiments. Also note that, on this understanding, theory-ladenness does not mean endorsement: It just means that an experiment cannot be understood without the theory in question, even if the experiment provides evidence against the theory – as Michelson and Morley's study did with respect to the Ether Wind Hypothesis.

In many cases, theoretical presuppositions concern essential features of a given phenomenon. Examples are the assumption that water is an element, that electromagnetic waves require luminiferous ether for transmission, that heat is a sub-





stance (e.g., caloric), or that consciousness is a cognitive phenomenon. While some of the traditional presuppositions (e.g., that water is one of the four elements) may be grounded in speculation or religious belief, scientifically relevant presuppositions today most typically are generalizations grounded in scientific observation. This holds for the assumption that heat is a substance, that electromagnetic waves require ether as a medium, or that consciousness is a cognitive phenomenon, as we will see below. As a consequence, theoretical presuppositions can be substantially modified or even dismissed upon further research – like the assumption that water is an element, that light requires ether, or that heat is a substance.

Note that our claim is not only that theoretical presuppositions actually play an important role in current neuroscientific research. Rather, we argue that they *need* to play this role because they help to make reasonable decisions based on the best available knowledge, e.g., regarding experimental design, the selection of hypotheses, and the interpretation of data.

Our second question concerns the theories with respect to which NCC research is supposed to be neutral. Let's first note that empirical research often requires the acceptance of very basic theoretical assumptions. For example, it doesn't make much sense for consciousness science to investigate the brain, if one does not accept the assumption that, in some way or other, conscious experience is based on brain activity. Likewise, you probably accept the BOLD-effect as a proxy for neural activity, including the underlying theories (Logothetis et al., 2001), if you run fMRI experiments. We assume that Koch and Chalmers accept the compatibility of this very basic level of theory-ladenness with their idea of theory-neutrality. This might be one of the reasons why Chalmers sometimes relativizes his neutrality claim.

As far as the remaining theoretical commitments are concerned, we suggest a threefold distinction between (1) a metaphysical, (2) an epistemological and (3) a neuroscientific level of theory-ladenness.

(1) Theories on the metaphysical level deal with the identity or non-identity of the mental and the physical. Understood in this way, theory-neutrality would sidestep the metaphysical question whether mental states are physical states. While neither Chalmers nor Crick and Koch are very explicit regarding this issue, the definition of NCCs they embark on does not seem completely neutral in this respect, though. Their definition is incompatible particularly with current forms of interactive dualism as proposed e.g., by Eccles or Libet. Both authors assume that consciousness requires some sort of non-physical activity, e.g., by non-physical "psychons" (Eccles, 1994) or "conscious mental fields" (Libet, 1994; Libet, 1996). This means that, contrary to one of the basic assumptions of the NCC program, neural activities cannot be "minimally sufficient" for conscious experience.

However, this incompatibility could be easily removed by minor adjustments of the NCC definition. These adjustments can make room for non-neural conditions,

and there are reasons to do so because this move would allow researchers with different metaphysical beliefs to participate in NCC studies. In fact, both Eccles and Libet were neuroscientists who conceded that – their dualist commitments notwithstanding – brain activities are essential for the emergence of consciousness; Chalmers (1998, 2000) explicitly refers to Libet’s NCC research. This is one of the reasons why such an adjustment of the original definition seems suggestive. Let’s conclude then, that NCC research is thought to be (or could easily be made) neutral regarding metaphysical questions, even if this is not reflected by the “official” definition.

(2) Theories on the epistemological level deal with explanations of phenomenal experience. Understood in this way, the most important implication of theory-neutrality would be that NCC research sidesteps issues like the Hard Problem of Consciousness. Apparently, this is a likely candidate for the understanding of theoretical neutrality. Koch and Crick’s (2002) three-step model, that has already been mentioned above, explicitly distinguishes between the search for the NCCs as the first step, the investigation of the causal link between neural activities and subjective feelings as a second step and, finally, the Hard Problem as something that future research may take care of in a third step.

A similar view can be found in Chalmers who explicitly distinguishes correlation from explanation and stresses that “it certainly does not follow that an NCC will yield an explanation of consciousness” (Chalmers, 2000). As already mentioned above, Chalmers, like Crick and Koch, envisages a three-step development in consciousness research. This implies a clear distinction between the NCC program as a first step, the development of functional explanations as a second step, and the solution of the Hard Problem as a third step. This step is taken as a remote possibility that might come up even later *after* the completion of the search for functional explanations.

“But who knows: somewhere along the line we may be led to the relevant insights that show why the link is there, and the hard problem may then be solved” (Chalmers, 1998, p. 25).

In a similar vein, Chalmers (2000) distinguishes between “theoretically neutral” NCC research and an explanation of consciousness which is understood as a further step:

“Once we have found an NCC, one might hope that [...] it will turn out to yield an explanation of consciousness, but these are further questions” (Chalmers, 2000, p. 37).

All this shows that both Chalmers and Koch assume that NCC research is neutral with respect to a full explanation of consciousness, that is a solution of the Hard Problem which is regarded as an independent project that may be taken care of

at some point in the future. This means that the search for the NCCs can be pursued no matter whether or not they contribute anything to our understanding of consciousness and, conversely, no matter what your theoretical commitments regarding the explanation of consciousness might be. That is, these commitments will not be among the theoretical presuppositions guiding NCC research.

(3) But what about the third candidate? Does theoretical neutrality extend to the level of neuroscientific theories of consciousness like the Global Workspace Theory, according to Koch and Chalmers? This is indeed the case. Both Koch and Chalmers assume that neuroscientific theories of consciousness go beyond the purely correlational research that the NCC program focuses on. If we have another look at Koch's three-step model with NCC research as the first, an investigation of the causal link between consciousness and neural activity as a second, and a solution of the Hard Problem as a third step, it seems obvious that neuroscientific theories of consciousness which actually try to establish causal links between specific neural activities and conscious experience belong to this second step. Likewise, Chalmers distinguishes between, first, the identification of the NCCs, second, "mechanistic explanations" of the functional properties of consciousness and, third, a "full explanation of consciousness" that solves the Hard Problem. In Chalmers's case, the assignment of neuroscientific theories to the second step is even more obvious. Chalmers's paradigm for a relevant explanandum is global availability, taken from Baars's Global Workspace Theory. One of today's most important neuroscientific theories of consciousness, Dehaene's Neuronal Workspace Theory which has been derived from Baars's Global Workspace, does exactly this: It "explains global availability in the brain" (Chalmers, 1998).

Let's conclude then that the theory-neutrality claim is intended to distinguish research collecting the neural correlates of consciousness (step 1) from theories trying to *explain* consciousness either on a functional/neuroscientific (step 2) or on a phenomenal level (step 3). This interpretation complies with the familiar textbook distinction between correlates and causal explanations, thus helping us to understand why Koch and Chalmers chose the term "correlates" in the first place – even if there are numerous fringe cases between mere correlates and full-blown explanations. These fringe cases provide another reason why Chalmers slightly<sup>1</sup> relativizes the neutrality claim (Chalmers, 2000, p. 37) – even if he insists on a "large extent" of neutrality. In fact, his reference to the "standard usage" of the term "correlate" but even more so the explicit distinction between theoretically neutral correlational research in the NCC program and "theoretically loaded" (Chalmers, 2000, p. 38) explanations in the later steps makes it clear that theory-neutrality is what sets the NCC program apart from the following stages of consciousness research.

---

<sup>1</sup>The qualification is not always included, though: see Chalmers (2000, p. 38).





Also note that there is a tradeoff here: Of course, a weak interpretation of the neutrality claim, e.g. one that includes metaphysical neutrality only, could make the NCC program less susceptible for objections. The obvious downside, however, is that it would make the NCC program futile because we might end up with what neuroscientists do anyway. Conversely, a strong interpretation would help the NCC program to make a real difference but could make it more susceptible for objections. We will come back to this question below.

### 3 The case of water and H<sub>2</sub>O

So, let's keep in mind that the original NCC program tried to sidestep theoretical commitments particularly regarding the function and explanation of consciousness. It did so in order to make studies on the neural correlates of consciousness open to researchers with widely diverging theoretical commitments. Plausible as this program sounds, it stands in stark contrast to a widespread consensus in the philosophy of science according to which there are strong and mutual connections between theoretical commitments and empirical research.

As is fairly well known, epistemologists and philosophers of science like Duhem (1991), Kuhn (1962), Sellars (1956), and Quine (1981) have long since argued that empirical work in general depends on theoretical commitments. But what, exactly, does this general epistemological principle of theory dependence mean when the identity of an everyday phenomenon like pain and a scientific kind, e.g., a specific type of neural activity, is at issue? And is it really true that theory dependence as postulated by Duhem, Kuhn, Sellars, and Quine is incompatible with the theory-neutrality claim of the NCC program? In order to answer these questions, we will discuss an example that has played a pivotal role in the consciousness debate: The identification of water and H<sub>2</sub>O which has been taken as a paradigm for an identity claim that is well-understood. That is why knowledge about the behavior of H<sub>2</sub>O molecules below 0° C can serve to explain the freezing of water. But how did scientists discover this relationship, and what can we learn from this discovery for the science of consciousness and, more precisely, for the questions above?

Here want to show, first, that the process that led to the identification of water and H<sub>2</sub>O was not only driven by current experimental results, but also by competing theoretical presuppositions. Second, we want to demonstrate that the relationship was mutual. While theoretical presuppositions were instrumental for guiding empirical research, they could also be modified and even dismissed based on empirical evidence, at least in the long run.

Our first point is that the water/H<sub>2</sub>O identity claim was not merely a product of empirical evidence. Rather, as Chang (2012) has shown, the identity claim came out as the result of the competition between two theories, the Phlogiston- and the Oxygen Theory which guided the design of experiments and the interpretation of

the results. Moreover, the competition was not decided by empirical evidence only but largely by pragmatic reasons.

As the competing theories led to opposing theoretical presuppositions, the case at hand can also help to illustrate what these presuppositions are and how they work. The presupposition of the Phlogiston Theory defended by Priestley and Ritter was the traditional idea that water is an element that does *not* allow for further decomposition. By contrast, Lavoisier and the proponents of the Oxygen Theory argued that water is a compound that *can* be decomposed into oxygen and hydrogen. As we will see, these presuppositions did not only guide the design of the experiments, and the research questions, they also affected how the results were interpreted, and, eventually, they were decisive for the outcome of the debate (Chang, 2012).

For example, the famous Electrolysis Experiment by Nicholson and Carlisle was obviously motivated by the claim that water is a compound of oxygen and hydrogen. Immediately before conducting this experiment, Nicholson and Carlisle had tried to construct a battery following Volta's description. In order to improve the contact between the battery and the cables attached to it, they put a few drops of water on the connection. But due to a design failure, there was no direct contact such that cable and battery worked as anode and cathode and thus decomposed water into hydrogen and oxygen, which was realized by the experimenters.

And this is where the theoretical presupposition comes in: Since Nicholson and Carlisle knew that this point was critical for the debate between the Oxygen and the Phlogiston Theory, they decided to do another experiment in order to test the decomposition of water into hydrogen and oxygen (Chang, 2012). The success of this test seemed to provide support for the Oxygen Theory and its presupposition that water was a compound of oxygen and hydrogen. But rather than giving in, the proponents of the Phlogiston Theory came up with an interpretation based on *their* presupposition, namely that water is an element.

Eventually though, the Oxygen Theory prevailed, but, as Chang shows, the reason was not support by converging evidence. The proponents of the Oxygen Theory were even unable for quite some time to overcome a fundamental objection that had been raised by Phlogiston theorists like Ritter, the so-called Distance Problem, which was not solved before the next turn of the century: How can oxygen and hydrogen move from one and the same molecule to two completely different locations, namely the anode and the cathode (Chang, 2012, p. 73 sq.)? But if the Oxygen Theory and the water/H<sub>2</sub>O identity claim were accepted already four decades earlier, this was mainly for pragmatic reasons: The Oxygen Theory was a better fit for Lavoisier's analytic chemistry since it was based on the same theoretical presupposition, namely that even basic substances could turn out to be compounds.

By contrast, the competing presupposition that water was an element was eventually dismissed. This reveals another important aspect of the relation between empirical studies and theoretical presuppositions: Even if presuppositions

guide the design and interpretation of individual experiments, they may be affected by the outcome of these experiments themselves, at least in the long run. This can happen, if there is a substantial amount of converging evidence that can be easily interpreted by (or strongly supports) one specific presupposition but not by a competing one. We will come back to this point below in Part 5.

## 4 The dependence of theories and data in consciousness research

So let us conclude, first, that the identification of water and H<sub>2</sub>O did not emerge directly from the evidence available, rather it came as a result of the competition between two opposing theories which provided different theoretical presuppositions for empirical research. And second, the success of the Oxygen Theory and the defeat of the Phlogiston Theory had a strong effect on the respective theoretical presuppositions as well.

This shows that the relation between theoretical presuppositions and empirical evidence is mutual. First, the search for evidence is guided by theoretical presuppositions. Second, the evidence, in turn, may affect our theoretical presuppositions. Thus, empirical research does not simply produce data that help to decide a pre-defined question about a given phenomenon, rather it results in changing our understanding of the phenomenon itself by way of changing or dismissing even basic assumptions about it, that is, theoretical presuppositions. If any of these observations holds for NCC research as well, then it is highly unlikely that this research can be theory-neutral.

But do these observations really hold for consciousness research? In what follows, we will argue that they do. Contrary to what the original idea of the NCC program suggests, finding the NCCs is not, and cannot be, theory-neutral in the sense mentioned above. This means that very basic assumptions about the nature of consciousness which we have called theoretical presuppositions are needed in order to make progress in empirical research – very much like these principles were needed in the process of determining the “chemical correlate of water”. And as it was the case in the water/H<sub>2</sub>O debate, empirical research and theoretical presuppositions are mutually related in NCC research as well: On the one hand, this research is guided by certain basic presuppositions regarding the nature of consciousness or some particular kind of conscious experience like pain. These presuppositions most typically come from one of the current theories about consciousness or some of its subfunctions (e.g. pain) and they guide experimental design, the generation of hypotheses, and the interpretation of the data. On the other hand, these presuppositions, in turn, are up for confirmation or revision once new empirical findings come in. They may thus lead to a further development of the underlying theories and help to improve our understanding of consciousness.



#### 4.1 How theoretical presuppositions guide consciousness research: The debate between cognitive and non-cognitive theories of consciousness

One of the probably best examples for how theoretical presuppositions affect NCC research is the debate about Cognitive and Non-Cognitive Theories of Consciousness (Overgaard & Grünbaum, 2012). The debate has pitted Cognitive Theories like the Global/Neural Workspace Theory by Baars (1996, 1997) and, later, Dehaene (2006), or Higher Order Theories of Consciousness by Rosenthal (2011; 1986) against Non-Cognitive approaches like Block's account of Access- and Phenomenal Consciousness (Block, 1995) or Lamme's Recurrent Loop Theory (Lamme, 2006). Here we will focus on the Global Workspace Theory as a paradigmatic cognitive approach and on Block's account as an example for a non-cognitive approach.

We take this debate to illustrate that the dependence between theories and experimental investigation is essential for NCC research as well. For example, we need a basic, theoretically grounded, idea of what "consciousness" means and what its basic functions are if we want to design NCC experiments or draw conclusions from them regarding the NCC. In other words: We need theoretical presuppositions if we want to collect and interpret new data. Moreover, it would be negligent to ignore the current state of knowledge and to refrain from trying to use and apply this knowledge as good as possible in the design and interpretation of empirical studies.

As both Dehaene's and Block's approaches are fairly well known, we will focus on the most relevant differences between them and will then try to make clear how the underlying understanding of consciousness affects experimental work on the NCCs.

Dehaene's Neuronal Workspace Theory goes back to Bernard Baars's Global Workspace Theory which is based on a fundamental assumption about what consciousness is and what it does:

"Consciousness seems to be the publicity organ of the brain. It is a facility for accessing, disseminating and exchanging information, and for exercising global coordination and control" (Baars, 1997, p. 299).

This understanding of consciousness as the "publicity organ of the brain" is a clear theoretical presupposition, which has an important consequence, namely that consciousness is closely connected to reportability. Let's call this the *Consciousness/Reportability Presupposition*: If I have a conscious pain experience and consciousness is the publicity organ of the brain, then I should be able to report this experience. This presupposition has guided Dehaene's research on the neural correlates of consciousness for many years (Dehaene & Naccache, 2001). In the present context, this theoretical presupposition has two important implications for empirical studies.

First and methodologically speaking, if consciousness is closely connected to reportability, then verbal report should be the gold standard for scientific access to consciousness in experimental design. In fact, this is exactly what Dehaene claims explicitly:

“Subjective reports are the key phenomena that a cognitive neuroscience of consciousness purports to study. As such, they constitute primary data that need to be measured and recorded along with other psychophysiological observations” (Dehaene & Naccache, 2001, p. 3).

Dehaene, however, has not only postulated that reportability is a key ingredient of consciousness, rather, against Block’s conceptual distinction of Access and Phenomenal Consciousness, he argues that there is even a conceptual connection between phenomenal consciousness and reportability:

“If the only support for the existence of phenomenal consciousness comes from conscious reports, then we find no reason to accept a major distinction between these two concepts” (Naccache & Dehaene, 2007, p. 519).

Second, Dehaene’s consciousness/reportability presupposition directly affects the interpretation of the data in NCC experiments. If reportability is a basic ingredient of consciousness, then the neural correlates of the cognitive processes underlying this reportability which are located primarily in the PFC count as neural correlates of consciousness. And this is, in fact, another main claim of the Global Workspace Theory:

“Neurophysiological, anatomical, and brain-imaging data strongly argue for a major role of prefrontal cortex, anterior cingulate, and the areas that connect to them, in creating the postulated brain-scale workspace” (Dehaene & Naccache, 2001, p. 1, sq.).

Block, by contrast, has developed quite a different view about the nature of consciousness which leads him to a rejection of the consciousness/reportability presupposition. Block distinguishes between two sorts of consciousness: *Access and Phenomenal Consciousness* (Block, 1995). While Block’s concept of *Access Consciousness* is well compatible both with the Global Workspace Theory<sup>2</sup> (Block, 2005) and the consciousness/reportability presupposition, it is his idea of *Phenomenal Consciousness* that leads him to reject this presupposition. In Block’s view, a person may well have a specific phenomenal experience without being able to report it. Phenomenal experience and, therefore, consciousness may “overflow” reportability (Block, 2007).

---

<sup>2</sup>In fact, Block explicitly endorses the Global Workspace Theory, but only as far as access consciousness is concerned: “Access conscious content is content information about which is ‘broadcast’ in the global workspace” (Block, 2005).



The rejection of the consciousness/reportability presupposition does not only set Block's account of Phenomenal Consciousness apart from the Global/Neuronal Workspace Theory, it can also serve to illustrate the consequences that this presupposition has both for experimental design in NCC research and for the interpretation of the data regarding possible NCCs themselves. First, as for experimental design, Block opposes Dehaene's reliance on report-based methods since – in Block's view – these methods are unable, in principle, to detect phenomenal experiences that come in the absence of reportability or access consciousness:

“To insist on introspective reportability as the gold standard is to encourage leaving out cases in which subjects have experiences without higher order thoughts” (Block, 2005, p. 50).

As a consequence, Block insists on methods that are able to provide evidence for conscious experience independent of subjective reports. One promising approach is provided by so called “no report paradigms”, which are able to detect conscious experience without asking for verbal report. We will discuss this method and its implications for the consciousness/reportability presupposition below in Part 5.

Second, as far as the interpretation of experimental data regarding possible NCCs is concerned, it would follow that neural activities resulting from verbal reports and the underlying cognitive processes do not count as NCCs. Rather, they count as confounds (Michel, 2017). Moreover, because Block assumes two distinct sorts of consciousness, he has to expect two different types of NCCs as well: Phenomenal and Access NCCs. While Block, like Dehaene, assumes that the NCCs of access consciousness can be found in frontal and parietal areas, phenomenal NCCs are most likely located in peripheral areas, e.g., in the occipital lobe if visual experiences are concerned (Block, 2005).

These observations have further consequences: Theoretical presuppositions are required even in NCC research both regarding the methodology and the interpretation of empirical results. Neutrality is not an option here: Methodologically speaking, we have to decide whether or not we accept self-reports as the gold-standard in consciousness research. And as far as the interpretation of the results are concerned, we must determine whether we take report-based neural activity as part of an NCC or as a confound. Both decisions depend on fundamental theoretical presuppositions regarding the nature of consciousness. Obviously, these decisions should be made on the basis of the best knowledge available. And this is what theoretical presuppositions help us to do.

## 4.2 Chalmers's pre-experimental bridging principles

It could be argued, though, that Chalmers does not deny the existence of theoretical presuppositions. He does not use this term, but it might be argued that what he calls *Pre-Experimental Bridging Principles* is basically the same as what we call *theoretical presuppositions*. Let's first try to make clear what Chalmers has in mind

when he talks about *Pre-Experimental Bridging Principles* and then discuss their relationship to *theoretical presuppositions*.

Chalmers thinks there is an obvious reason why bridging principles are needed in NCC research: We lack direct access to others' conscious experience. Bridging principles step in and allow us to assign conscious experience to experimental subjects on the basis of proxies like verbal reports or behavioral data. There are two such bridging principles: First, the *Principle of Verbal Report* which holds that a mental state is conscious if it is subject to verbal report. The second principle holds that a mental state is conscious if it is directly available for global behavioral control. Let's call it the *Principle of Behavioral Control*.

Chalmers calls these principles "pre-experimental" because he thinks they do not depend on scientific theories: That is why they are compatible with the theory-neutrality claim. Rather, they are based on an analysis of the concept of CONSCIOUSNESS and on first-person experience, so they neither call for empirical support nor can empirical evidence lead to substantial revision or improvement of these principles (Chalmers, 1998). This is an obvious difference between Chalmers's bridging principles and theoretical presuppositions, because the latter, unlike the former, are generalizations from empirically supported theories. This is why bridging principles are compatible with the demand for theory-neutrality while theoretical presuppositions are not. And since theoretical presuppositions are supported by empirical evidence, they can not only be refined – as Chalmers seems to concede for the bridging principles<sup>3</sup> –, rather, they can be substantially revised or even dismissed entirely if a sufficient amount of new evidence comes in that is incompatible with a given presupposition. We have shown this above with respect to the debate between the Phlogiston and the Oxygen Theory and we will provide further evidence from NCC research below in Part 5. There, we demonstrate in some detail how empirical results may affect theoretical presuppositions like the consciousness/reportability presupposition.

Finally, Chalmers's pre-experimental principles are dedicated to *one* very specific issue: the interpretation of verbal reports and behavior with respect to consciousness. But, again, even if the content of the consciousness/reportability presupposition is somewhat similar to Chalmers's principle of verbal report: It is just *one* of an infinitely large number of potential theoretical presuppositions which may cover all kinds of subjects, e.g. the regions of interest underlying the experience of pain, the role of the Amygdala in negative emotions, or the role of working memory in conscious experience: Almost any knowledge that may appear helpful in the search for the neural correlates of consciousness can be used as a theoretical presupposition.

---

<sup>3</sup>Chalmers seems to concede a limited amount of refinement, particularly through first-person experiments, but his position here is unclear. Initially he seems to exclude refinements through normal third person experiments altogether, claiming that we "can't refine them [i.e. the pre-experimental principles] experimentally (except perhaps through first-person experimentation)" (Chalmers, 1998, p. 3). Later he concedes that these "principles may themselves be refined as experimental research goes along" (Chalmers, 1998, p. 6).

In sum, we think that, for all these reasons, Chalmers's understanding of pre-experimental bridging principles is significantly different from what we have described as theoretical presuppositions and thus provides another piece of evidence showing that Chalmers does not even implicitly accept theoretical presuppositions.

Let us add as an aside that there are reasons to believe that Chalmers's two bridging principles are not pre-experimental. In other words: Experimental evidence *does* speak to the principle of verbal report and to the principle of behavioral control. First, there is some evidence which might be seen to challenge the principle of verbal report. Nisbett and Wilson (1977) or Gazzaniga and Ledoux (1978) have shown already in the 1970s that, under certain conditions, subjects tend to report mental states that they did not experience consciously, e.g., when they would otherwise lack sufficient explanation for their behavior. An even more striking example comes from patients suffering from Anton's Syndrome (Anton, 1899; Maddula et al., 2009; Kini & Khatker, 2019), a rare case of visual anosognosia. Anton's patients are actually blind, but they claim they have visual perceptions even if there is strong evidence against these claims. Thus, contrary to the principle of verbal report, neurologists take these claims as confabulations (Carvajal et al., 2012; Das & Naqvi, 2021). Moreover, contrary to the principle of behavioral control, priming experiments show that our behavior can well be affected by pieces of information that escape conscious experience (Kristjánsson & Campana, 2010).

Note again that we do not argue for a rejection of Chalmers's bridging principles. Rather, we want to show that – contrary to Chalmers – empirical evidence does speak to these principles, e.g., because it can help us to specify the conditions under which these principles hold, but it may also force us to dismiss these principles. All this gives us strong reasons to reject Chalmers's claim that the two principles he mentions are pre-experimental. And as the pre-experimental status of the bridging principles was essential for their compatibility with the theory-neutrality claim of the NCC program, Chalmers seems to be forced to make a choice: Either, he has to give up the bridging principles in order to save the neutrality claim – this however would make it difficult for NCC research to assign consciousness on the basis of proxies. Or he has to give up the theory-neutrality claim, which would jeopardize an essential feature of the NCC program.

Since the above considerations are decisive for our claim that NCC research cannot be theory-neutral, a brief intermediate summary might be in place. So, let's ask flat out why can't Chalmers and Koch simply accept the observations made above? Why can't they insist that theory-neutrality just means that NCC research sets metaphysical questions and maybe questions related to the Hard Problem aside, while allowing for theoretical presuppositions resulting from neuroscientific research?

There are mainly two reasons why this is not an option for Koch and, particularly, for Chalmers. First, we have shown that NCC research, as any other empirical investigation, needs theoretical presuppositions in order to make unavoid-

able decisions about research questions, experimental design, and the interpretation of the data. But if you accept theoretical presuppositions like the consciousness/verbal report presupposition, then you make a commitment regarding the underlying theory, e.g. the Global Workspace Theory that justifies the presupposition. We think that already this commitment, taken by itself, is a clear violation of theory-neutrality. Moreover, theories of consciousness like the Global Workspace Theory are explicitly assigned both by Chalmers and by Koch to the second, “theoretically loaded” stage of consciousness research. That is, these approaches do not count as theory-neutral in Chalmers’s and Koch’s own view because research in this second phase is supposed to provide functional or mechanistic explanations of consciousness, rather than mere correlations – which is exactly what the Global Workspace Theory does. Finally, we have seen that Chalmers’s plea for pre-experimental bridging principles cannot be read as an implicit acceptance of theoretical presuppositions. Quite the contrary: Chalmers goes to a lot of trouble to show that his bridging principles do not imply any theoretical commitments – as our theoretical presuppositions do. Bridging principles are thought to follow from conceptual considerations and first-person experience, but not from experimentally based theories and thus are supposed to be compatible with the claim for theory-neutrality. But as we have seen, there are strong reasons to doubt that this is actually the case for the two principles that Chalmers identifies.

There is still a second, independent reason, why accepting theoretical presuppositions or similar kinds of theoretical commitments is not an option for Koch and Chalmers. Both authors seem to assume that the NCC program really makes a difference compared to the practice in contemporary consciousness research. Otherwise, the NCC program would just preach to the converted because it would “demand” what neuroscientists are doing anyway. That is, contrary to what Koch and Chalmers claim, the NCC program would not help to attract new scientists to the field of consciousness studies, irrespective of their theoretical commitments. For all these reasons, a weak reading of the theory-neutrality claim that would be compatible with the above considerations is not available for Koch and Chalmers.

## 5 How empirical results may affect theoretical presuppositions

So far, we have argued that – against the claim of theory-neutrality – NCC research like empirical science in general is informed by theoretical presuppositions about the nature of consciousness and it needs to be so because these assumptions guide inevitable choices researchers have to make regarding experimental design, research questions, and the interpretation of their data.

In this last part of our paper, we want to show that the relationship between empirical research and theoretical presuppositions is mutual: Empirical research, that is, does affect theoretical presuppositions. Moreover, we will demonstrate,



that this process – if successful – can help to improve our understanding of the phenomena under investigation.

In order to show this, we will present two examples that illustrate two slightly different ways of how empirical evidence may affect theoretical presuppositions. In the first section, we present the discussion on *No-Report Paradigms* in order to demonstrate how empirical studies may help to resolve conflicts between opposing theoretical presuppositions. This is not only a technical or methodological point. Rather, given that theoretical presuppositions concern basic features of a phenomenon, dismissing a presupposition that turns out to be unfounded may help to improve our understanding of that phenomenon. Our second example will be pain research and pain measurement. We will show how empirical findings can help to generate new theoretical presuppositions, which can then be used in subsequent studies. And again, this may help to improve our understanding of the phenomenon as well.

While we don't want to suggest that Chalmers's understanding of theoretical neutrality is incompatible in general with the possibility that NCC studies have theoretical consequences, we think that a discussion of the effects empirical studies have on theoretical presuppositions is essential for a full understanding of theoretical presuppositions. This holds in particular for our understanding how these presuppositions emerge and how we resolve conflicts between them – even if this requires that we take a look also at the consequences of NCC research.

## 5.1 No-report paradigms

The general assumption underlying Chalmers's two pre-experimental bridging principles is the idea that global availability can be taken as a proxy for consciousness: "When information is *directly available for global control* in a cognitive system, then it is conscious" (Chalmers, 1998). But this assumption is not *a priori* at all but, rather, it is one of the basic tenets of the Global Workspace Theory of consciousness – as we have already seen above. In fact, Chalmers (1998) explicitly refers to the Global Workspace Theory whose proponents insist that their theory "is based entirely on well-established empirical contrasts between pairs of conscious and unconscious events" (Baars, 1997; Michel & Morales, 2020). In doing so they confirm our point that experimental data do speak to those allegedly "pre-experimental" principles.

More recently, the principle of global availability has become the subject of a controversial debate about so-called "no-report paradigms" in consciousness studies which is directly related to the debate between cognitive and non-cognitive theories of consciousness mentioned above (Block, 2019; Frässle et al., 2014; Michel, 2017; Michel & Morales, 2020; Naber et al., 2011; Tsuchiya et al., 2015). As their name indicates, no-report paradigms try to provide evidence for conscious experience without relying on first-person reports.

This is why no-report paradigms speak to the Global Workspace Theory's basic tenet, according to which global availability is essential for consciousness, a claim



that has often been used as a theoretical presupposition: If the Global Workspace Theory were right, we would expect that conscious experience cannot be dissociated from global availability. Thus, we should find activity in the PFC where the “publicity organ” of the brain is supposed to be located, whenever a participant has a conscious experience, even if they are not requested to report. By contrast, if non-cognitive theories were right, we would expect no-report paradigms to provide evidence for conscious experience without reportability, and evidence for activity in the peripheral areas in the absence of PFC activity and global availability. This would support Block’s claim that phenomenal and access consciousness should be distinguished. Previous evidence for PFC activity from report-based paradigms could then be attributed to a “methodological artefact” (Michel, 2017) due to confounds in report paradigms. Note that this discussion provides another piece of evidence that Chalmers’s principles, no matter whether or not they turn out to be valid, may not be pre-experimental.

One no-report paradigm has emerged from binocular rivalry-experiments (Frässle et al., 2014; Naber et al., 2011). In a typical design, two different optical stimuli moving into opposite directions are projected to the left and the right eye, e.g., a green grid moving to the right is projected to the right eye and a red grid moving to the left to the left eye. The visual system will then switch between the two stimuli every now and then such that, at any given point in time, either the green grid moving to the right or the red grid moving to the left will be dominant in the participant’s experience. Normally, participants report which stimulus is dominant and when the switch occurs.

Frässle et al. (2014) however used the *optokinetic nystagmus* instead to determine the dominant stimulus. The optokinetic nystagmus is an involuntary eye-movement that follows moving visual stimuli, but also reveals the dominant stimulus in a binocular rivalry experiment.

Using this paradigm for a comparison between the neural correlates in a report and a no-report condition, Frässle et al. could show that the PFC was active in the report condition but not in the no-report condition, although conscious experience did not differ between these two conditions. As essential functions underlying reports (working memory, attention) are located in the PFC, the authors concluded that PFC activity, contrary to what the Global Workspace Theory claims, is *not* a correlate of conscious experience itself but, rather, just a correlate of introspective reporting.

Unsurprisingly though, proponents of the Global Workspace Theory reject this interpretation. They see evidence for PFC related activity even in no-report paradigms like Frässle et al.’s study (Michel & Morales, 2020). Other authors have tried to show that even more sophisticated “no-cognition paradigms” (Block, 2019) are unable, in principle, to rule out the existence of neural correlates of consciousness in the PFC, as proposed by the Global Workspace Theory (Phillips & Morales, 2020).

Again, we will not take sides in this debate, nor will we make any commitment as to whether or not PFC activity is an NCC. Rather, our point is that the

discussion about no-report paradigms provides further evidence, that the dependence between empirical evidence and theoretical commitments is mutual – just as it turned out to be in the water/H<sub>2</sub>O case.

Here we see that empirical evidence from no-report paradigms speaks to the debate whether or not global availability is an essential feature of consciousness, that is, whether or not we should accept the consciousness/reportability presupposition mentioned above, or Chalmers's – allegedly pre-experimental – principle of verbal report. That is, theoretical presuppositions, even if they are quite abstract, are subject to empirical confirmation and disconfirmation. As these presuppositions most typically express basic assumptions about the phenomenon of consciousness, this continuous process of testing and revisioning of our basic assumptions concerning consciousness should lead to an improved understanding of the phenomenon, at least in the long run. In the case at hand, we should expect a better understanding of the relationship between conscious experience and reportability.

## 5.2 Pain research

Our second example for the mutual relationship between NCC research and theoretical presuppositions is taken from recent developments in pain research. While the example above has illustrated how empirical data can affect theoretical presuppositions, we will now go one step further and show how theoretical presuppositions that have emerged from empirical research can guide further studies. More specifically, we use pain research as an example that shows more precisely how empirical research on pain can help to develop new theoretical presuppositions, and how these presuppositions guide new experimental work.

Pain experience is one of the paradigmatic issues in the debate on consciousness since the nineteen-fifties both in philosophy and neuroscience, and it has been an important research subject ever since, also due to its clinical relevance. Recent research seems to indicate that the seemingly homogenous phenomenon we call pain can be distinguished further into two sub-components, namely *sensory-discriminative pain* on the one hand, and *affective pain* on the other (Gligorov, 2017; Gracely et al., 1982; Rainville et al., 1997). Both components can be differentiated from the first-person phenomenal perspective as well as from the third-person neuroscientific perspective. While sensory-discriminative pain lets you feel the kind and the place of the tissue damage and involves activity in the somatosensory cortex, affective pain is relevant for the aversive character of pain experience and is realized by networks involving the anterior cingulate cortex. The point at issue here is that this differentiation, apart from allowing for more accurate and valid measurements, gives us a more precise idea of the experience of pain and its neural substrate, thus making good on our claim that NCC research can improve our understanding of the phenomenon of pain.

### 5.3 Neural correlates of experienced and empathic pain

In order to show how this improvement plays out in empirical research, we will turn to a more recent study of the neural correlates of experienced and empathic pain. We will demonstrate how the distinction between affective and sensory pain, that is, one of the results from recent pain research, has then been used as a *theoretical presupposition* in a subsequent study, illustrating one further step in the mutual relationship between theoretical presuppositions and empirical research. Our example will be a study by Singer and colleagues (2004) that compares the neural correlates of one's own pain experience with the correlates of the empathic response to someone else's pain. The study uses also other theoretical presuppositions, e.g., for methodological reasons. Like many other pain-researchers (Gligorov, 2017; Makin, 2016; Woo et al., 2017) the authors rely on the so-called pain matrix (Legrain et al., 2011; Prichep et al., 2011) which specifies the regions where neural correlates of pain can be expected.

The important point, however, are the theoretical presuppositions Singer et al. use for the interpretation of their data. In a first step, they could show that anterior cingulate cortex and insula are active both in experienced and empathic pain, while the secondary somatosensory cortex is active in experienced pain only. Taken by itself, this finding does not tell us very much.

But the authors go on to interpret their correlational data with the help of a theoretical presupposition, namely the functional and phenomenal distinction between sensory-discriminative and affective pain, that is one of the theoretical consequences of the empirical studies mentioned above. Capitalizing on this presupposition, the authors can show that it is the affective aspect of pain, which is relevant for empathic pain. One's own pain experience, by contrast, also includes more detailed sensory-discriminative information about the location, quality and intensity of the tissue damage which is less relevant when we empathize with others.

Note that this interpretation is not just a contingent addendum. Rather the theoretical presuppositions about the function and phenomenology of the different areas underlying pain help to make sense of the data, they improve our understanding of pain in general and empathic pain in particular, and they open up new possibilities for behavioral tests: E.g. if Singer's interpretation is correct, we would expect that, compared to first-person pain experience, the epistemic disadvantage of empathic pain primarily concerns the location and quality of the other's tissue damage but much less so the affective character of the pain experience.

But why couldn't Chalmers insist that the above assumptions about the role of the ACC, the insula and the secondary somatosensory cortex are compatible with his neutrality claim? Most importantly, these assumptions are clearly functional claims which, according to Chalmers – and Koch – are reserved for the second stage of consciousness research. This is a significant difference since, as we have already seen, we have to wait “a century or two” until we arrive at this stage according to Chalmers. And the fact that these functional assumptions are “only” about

pain and not about consciousness in general can be easily explained by the focus of Singer et al.'s paper which is "only" on pain rather than on consciousness in general. This indicates, again not very surprisingly, that there might a correspondence between the scope of a study and the scope of the theoretical presuppositions the study makes use of.

## 6 Conclusion

Let's take stock. We have started with the original idea of the NCC program according to which finding the neural correlates of consciousness is an empirical endeavor that remains neutral regarding some of the main theoretical controversies about consciousness. The program suggests that we can accumulate empirical facts which are useful for all the participants in the debate, independently of their divergent theoretical commitments. Most importantly, the NCC program is intended to sidestep controversies regarding neuroscientific explanations of consciousness and the Hard Problem of Consciousness: Chalmers and Koch assign these two issues to subsequent "theoretically loaded" steps in the development of the neuroscience of consciousness, and these steps are clearly separated temporally and systematically from the "theoretically neutral" search for the neural correlates of consciousness as the first step.

Plausible as it sounds, the neutrality claim is in stark tension with widely accepted epistemological arguments brought forward by authors like Duhem, Quine, Kuhn, or Sellars who insist on a strong dependence between theoretical and conceptual commitments on the one hand, and empirical findings on the other. We tried to show that NCC research does not make an exception here. Like any other scientific endeavor, the search for the neural correlates of consciousness requires theoretical presuppositions which affect the methodology, help to figure out promising hypotheses and research questions, and support the interpretation of the results. The dependence is mutual: Our theoretical and conceptual commitments are affected by empirical results as well.

After demonstrating how the interaction between theoretical presuppositions and empirical research influenced the debate about the identity of water and H<sub>2</sub>O, we took the current debate between cognitive and non-cognitive theories of consciousness as well as pain research as examples for this mutual dependence. In the former debate, the decisive disagreement concerns the presupposition that first-person reportability is an essential component of consciousness. While cognitive theories accept this presupposition, non-cognitive approaches deny it. This has essential implications for the design of experiments and the interpretation of their data: Cognitive theories can accept first-person reports as an experimental method and the correlates of these reports as NCCs, while non-cognitive theories deny this. But the dependence between theoretical presuppositions and empirical evidence is mutual: Empirical results can and do affect our theoretical commitments as well. This means, first, that, at some point, converging evidence which is inconsistent



with a certain presupposition may lead us to revise or even dismiss this presupposition, like the claim that heat is a substance or that water is one of the four elements.

Taking pain research as our last example, we tried to show how the consequences of one series of experiments can become the presuppositions of subsequent studies. This mutual relationship between theoretical presuppositions and empirical studies which already concerns NCC research may result in a continuous process of theory testing and refinement which eventually improves our understanding of pain – or even of consciousness in general.

This is of particular importance because one of the most controversial issues the NCC program tried to sidestep, the Hard Problem, might be affected by an improved understanding of consciousness as well – even if it is impossible to predict when and how this will happen. But if empirical research enhances our understanding of the phenomenon of consciousness – as history witnesses and epistemological considerations show – NCC research may contribute to a solution of the Hard Problem.

If something like this is true, then the NCC program may help to demonstrate that consciousness is not a mystery. Rather, it is an ordinary – though extremely complex – problem of scientific study. And NCC research may, somewhat ironically, make a significant contribution to solving a problem that it was intended to sidestep.

### Acknowledgments

This research was supported by the German Research Foundation, grant RTG 2386, „Extrospection” and grant EXC 2002/1 ”Science of Intelligence”. I would like to thank Elisa Filevich, John-Dylan Haynes, Holger Lyre, Dimitri Coelho Mollo, Sera Schwarz, Ali Yasar, the members of the RTG Extrospection, and three anonymous reviewers for helpful suggestions and important comments.

## References

- Anton, G. (1899). Ueber die Selbstwahrnehmung der Herderkrankungen des Gehirns durch den Kranken bei Rindenblindheit und Bintaubheit. *Archiv für Psychiatrie Und Nervenkrankheiten*, 32(1), 86–127. <https://doi.org/10.1007/BF02126945>
- Baars, B. J. (1996). Understanding subjectivity: Global workspace theory and the resurrection of the observing self. *Journal of Consciousness Studies*, 3(3), 211–217.
- Baars, B. J. (1997). In the theatre of consciousness: Global workspace theory, a rigorous scientific theory of consciousness. *Journal of Consciousness Studies*, 4(4), 292–309.
- Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18(2), 227–247. <https://doi.org/10.1017/S0140525X00038188>
- Block, N. (2005). Two neural correlates of consciousness. *Trends in Cognitive Sciences*, 9(2), 46–52. <https://doi.org/10.1016/j.tics.2004.12.006>
- Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *The Behavioral and Brain Sciences*, 30(5-6), 481–548. <https://doi.org/10.1017/S0140525X07002786>
- Block, N. (2019). What is wrong with the no-report paradigm and how to fix it. *Trends in Cognitive Sciences*, 23(12), 1003–1013. <https://doi.org/10.1016/j.tics.2019.10.001>

Pauen, M. (2021). Why NCC research is not theory-neutral. *Philosophy and the Mind Sciences*, 2, 10. <https://doi.org/10.33735/phimisci.2021.9188>



©The author(s). <https://philosophymindscience.org> ISSN: 2699-0369



- Carvajal, J., Cárdenas, A., Pazmiño, G., & Herrera, P. (2012). Visual anosognosia (anton-babinski syndrome): Report of two cases associated with ischemic cerebrovascular disease. *Journal of Behavioral and Brain Science*, 2(3), 394–398. <https://doi.org/10.4236/jbbs.2012.23045>
- Chalmers, D. J. (1998). On the search for the neural correlate of consciousness. In S. R. Hameroff, A. W. Kaszniak, & A. C. Scott (Eds.), *Toward a Science of Consciousness II* (pp. 2–219). MIT Press.
- Chalmers, D. J. (2000). What is a neural correlate of consciousness? In T. Metzinger (Ed.) *Neural Correlates of Consciousness* (pp. 17–39). MIT Press.
- Chang, H. (2012). *Is Water H2O? Evidence, Pluralism and Realism*. Springer.
- Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences*, 2, 263–275.
- Das, M. J., & Naqvi, I. A. (2021). Anton syndrome [updated 2021 April 10]. In *StatPearls*, <https://www.ncbi.nlm.nih.gov/books/NBK538155/>.
- Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences*, 10(5), 204–211. <https://doi.org/10.1016/j.tics.2006.03.007>
- Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, 79(1), 1–37. [https://doi.org/10.1016/S0010-0277\(00\)00123-2](https://doi.org/10.1016/S0010-0277(00)00123-2)
- Duhem, P. M. M., Wiener, P. P., Vuillemin, J., & Broglie, L. de. (1991). *The Aim and Structure of Physical Theory*. Princeton University Press.
- Eccles, J. C. (1994). *How the Self Controls Its Brain*. Springer.
- Flohr, H. (1992). Qualia and brain processes. In A. Beckermann, H. Flohr, & J. Kim (Eds.), *Emergence or Reduction?: Essays on The Prospects of Nonreductive Physicalism* (pp. 220–238). W. De Gruyter.
- Frässle, S., Sommer, J., Jansen, A., Naber, M., & Einhäuser, W. (2014). Binocular rivalry: Frontal activity relates to introspection and action but not to perception. *Journal of Neuroscience*, 34(5), 1738–1747. <https://doi.org/10.1523/JNEUROSCI.4403-13.2014>
- Gazzaniga, M. S., & LeDoux, J. E. (1978). *The Integrated Mind*. Springer.
- Gligorov, N. (2017). Don't worry, this will only hurt a bit: The role of expectation and attention in pain intensity. *The Monist*, 100(4), 501–513. <https://doi.org/10.1093/monist/onx025>
- Gracely, R. H., Dubner, R., & McGrath, P. A. (1982). Fentanyl reduces the intensity of painful tooth pulp sensations: Controlling for detection of active drugs. *Anesthesia and Analgesia*, 61(9), 751–755.
- Haxby, J. V. (2012). Multivariate pattern analysis of fMRI: The early beginnings. *NeuroImage*, 62(2), 852–855. <https://doi.org/10.1016/j.neuroimage.2012.03.016>
- Haynes, J.-D. (2009). Decoding visual consciousness from human brain signals. *Trends in Cognitive Sciences*, 13(5), 194–202. <https://doi.org/10.1016/j.tics.2009.02.004>
- Kini, A., Khatker, M., Lee, A. G., Othman, B. A., & Ponce, C. P. (2019). Anton syndrome. *Eye Wiki: American Academy of Ophthalmology*. [https://eyewiki.aaopt.org/Anton\\_Syndrome](https://eyewiki.aaopt.org/Anton_Syndrome)
- Koch, C. (2004). *The Quest for Consciousness: A Neurobiological Approach*. Roberts & Co.
- Koch, C., & Crick, F. (2002). *The neuronal basis of visual consciousness*. <https://authors.library.caltech.edu/40600/1/465.pdf>
- Koch, C., Massimini, M., Boly, M., & Tononi, G. (2016). Neural correlates of consciousness: Progress and problems. *Nature Reviews Neuroscience*, 17(5), 307–321. <https://doi.org/10.1038/nrn.2016.22>
- Kristjánsson, Á., & Campana, G. (2010). Where perception meets memory: A review of repetition priming in visual search tasks. *Attention, Perception, & Psychophysics*, 72(1), 5–18. <https://doi.org/10.3758/APP.72.1.5>
- Kuhn, T. (1962). *The Structure of Scientific Revolutions*. University of Chicago Press.
- Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, 10(11), 494–501. <https://doi.org/10.1016/j.tics.2006.09.001>
- Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, 15(8), 365–373. <https://doi.org/10.1016/j.tics.2011.05.009>
- Legrain, V., Iannetti, G. D., Plaghki, L., & Mouraux, A. (2011). The pain matrix reloaded: A salience detection system for the body. *Progress in Neurobiology*, 93(1), 111–124. <https://doi.org/10.1016/j.pneurobio.2010.10.005>
- Libet, B. (1994). A testable field theory of mind-brain interaction. *Journal of Consciousness Studies*, 1, 119–126.
- Libet, B. W. (1996). Solutions to the hard problem of consciousness. *Journal of Consciousness Studies*, 3, 33–35.
- Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412(6843), 150–157. <https://doi.org/10.1038/35084005>
- Maddula, M., Lutton, S., & Keegan, B. (2009). Anton's syndrome due to cerebrovascular disease: A case report. *Journal of Medical Case Reports*, 3, 9028. <https://doi.org/10.1186/1752-1947-0003-0000009028>



- Makin, S. (2016). Imaging: Show me where it hurts. *Nature*, 535(7611), 8–9. <https://doi.org/10.1038/535S8a>
- Michel, M. (2017). Methodological artefacts in consciousness science. *Journal of Consciousness Studies*, 24(11-12), 94–117.
- Michel, M., & Morales, J. (2020). Minority reports: Consciousness and the prefrontal cortex. *Mind and Language*, 35(4), 493–513. <https://doi.org/10.1111/mila.12264>
- Michelson, A. A., & Morley, E. W. (1887). On the relative motion of the earth and the luminiferous ether. *American Journal of Science*, s3-34(203), 333–345. <https://doi.org/10.2475/ajs.s3-34.203.333>
- Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, 46(3), 774–785. <https://doi.org/10.1016/j.neuropsychologia.2007.10.005>
- Naber, M., Frässle, S., & Einhäuser, W. (2011). Perceptual rivalry: Reflexes reveal the gradual nature of visual awareness. *PLoS One*, 6(6), e20910. <https://doi.org/10.1371/journal.pone.0020910>
- Naccache, L., & Dehaene, S. (2007). Reportability and illusions of phenomenality in the light of the global neuronal workspace model. *Behavioral and Brain Sciences*, 30(5-6), 518–520. <https://doi.org/10.1017/S0140525X07002993>
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 231–259.
- Overgaard, M., & Grünbaum, T. (2012). Cognitive and non-cognitive conceptions of consciousness. *Trends in Cognitive Sciences*, 16(3), 137. <https://doi.org/10.1016/j.tics.2011.12.006>
- Phillips, I., & Morales, J. (2020). The fundamental problem with no-cognition paradigms. *Trends in Cognitive Sciences*, 24(3), 165–167.
- Prichep, L. S., John, E. R., Howard, B., Merkin, H., & Hiesiger, E. M. (2011). Evaluation of the pain matrix using EEG source localization: A feasibility study. *Pain Medicine (Malden, Mass.)*, 12(8), 1241–1248. <https://doi.org/10.1111/j.1526-4637.2011.01191.x>
- Quine, W. V. (1981). *Theories and Things*. Belknap Press.
- Rainville, P., Duncan, G. H., Price, D. D., Carrier, B., & Bushnell, M. C. (1997). Pain affect encoded in human anterior cingulate but not somatosensory cortex. *Science (New York, N.Y.)*, 277(5328), 968–971. <https://doi.org/10.1126/science.277.5328.968>
- Rosenthal, D. M. (1986). Two concepts of consciousness. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 49(3), 329–359. <http://www.jstor.org/stable/4319832>
- Sellars, W. (1956). Empiricism and the philosophy of mind. *Minnesota Studies in the Philosophy of Science*, 1(19), 253–329.
- Singer, T., Seymour, B., O’Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science (New York, N.Y.)*, 303(5661), 1157–1162. <https://doi.org/10.1126/science.1093535>
- Tsuchiya, N., Wilke, M., Frässle, S., & Lamme, V. A. F. (2015). No-report paradigms: Extracting the true neural correlates of consciousness. *Trends in Cognitive Sciences*, 19(12), 757–770. <https://doi.org/10.1016/j.tics.2015.10.002>
- Woo, C.-W., Schmidt, L., Krishnan, A., Jepma, M., Roy, M., Lindquist, M. A., Atlas, L. Y., & Wager, T. D. (2017). Quantifying cerebral contributions to pain beyond nociception. *Nature Communications*, 8(1), 14211. <https://doi.org/10.1038/ncomms14211>

## Open Access

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

