The psychological origins of the Hard Problem: How our consciousness is shaping the neuroscience of consciousness

Magnus Bendixen Englund

MAPSYK360, masterprogram i psykologi,

Studieretning: Atferd og nevrovitenskap

ved

UNIVERSITETET I BERGEN

DET PSYKOLOGISKE FAKULTET

VÅR 2023

Antall ord i hoveddelen: 19 913

Veileder: Mark C. Price, Institutt for samfunnspsykologi

Oppgaven er skrevet i henhold til formatteringen anbefalt i APA 7. utgave, med figurer og tabeller plassert i teksten.

Abstract

The field of consciousness studies contains a substantial number of coexisting neurocognitive theories of consciousness. These theories vary in their initial definitions of what consciousness is, biasing scientific methods and measurement of the phenomenon, resulting in a divided science. Definitions of consciousness tend to vary along two lines: Either consciousness is seemingly reducible to physical and functional processes, indicating what is called *access consciousness*; or it constitutes a seemingly irreducible experience, indicating *phenomenal consciousness*. These two types of definitions correspond to two opposing camps on what is referred to as *the hard problem of consciousness*, also called *the explanatory gap*. While much effort has been spent by each camp either criticizing or defending the hard problem, little work has been done to explain why the two camps vary along these lines. In other words, there is a gap between our positions on the explanatory gap, which I label the "meta-gap". In the current paper I contribute to bridging the meta-gap by attempting to explain and reconcile this basic disagreement in the field. By performing targeted literature searches, I answer seven research questions which serve as stepping stones to take us from problematic features of the field, to individual differences between researchers as a reason for these problems. My analysis of these individual differences results in two hypothesized psychological constructs: *Internal and external explanatory focus*. I conclude by indicating that solving the meta-gap involves becoming aware of our individual dispositions towards choosing different explanatory targets for consciousness.

Keywords: Theories of consciousness, hard problem, meta-problem, epistemology, individual differences, cognitive style

Sammendrag

Feltet bevissthetsstudier inneholder et betydelig antall sameksisterende nevrokognitive bevissthetsteorier. Disse teoriene varierer i deres innledende definisjoner av hva bevissthet er, noe som skaper skjevhet for vitenskapelige metoder og måling av fenomenet, som igjen resulterer i en splittet vitenskap. Definisjoner av bevissthet har en tendens til å variere langs to linjer: Enten er bevissthet tilsynelatende reduserbar til fysiske og funksjonelle prosesser, noe som indikerer det som kalles *tilgangsbevissthet*; eller så utgjør den en tilsynelatende ikke-reduserbar opplevelse, som indikerer *fenomenal bevissthet*. Disse to typene definisjoner svarer til to motstridende stillinger til det som omtales som *det vanskelige bevissthetsproblemet*, også kalt *forklaringsgapet*. Selv om hver leir har brukt mye krefter på enten å kritisere eller forsvare det vanskelige problemet, har det blitt gjort lite arbeid for å forklare hvorfor de to leirene varierer langs disse linjene. Det er med andre ord et gap mellom våre holdninger til forklaringsgapet, som jeg kaller "metagapet". I denne artikkelen bidrar jeg til å bygge bro over metagapet ved å forsøke å forklare og forsone denne grunnleggende uenigheten i feltet. Ved å utføre målrettede litteratursøk svarer jeg på syv forskningsspørsmål som fungerer som springbrett ved å ta oss fra problematiske trekk ved feltet, til individuelle forskjeller mellom forskere som årsaken til disse problemene. Min analyse av disse individuelle forskjellene resulterer i to antatte psykologiske konstrukter: *Internt og eksternt forklaringsfokus*. Jeg avslutter med å indikere at å løse metagapet innebærer å bli klar over våre individuelle disposisjoner for å velge ulike forklaringsmål for bevissthet.

Nøkkelord: Bevissthetsteorier, det vanskelige bevissthetsproblemet, metaproblemet, epistemologi, individuelle forskjeller, kognitiv stil

Acknowledgements

The idea for the current thesis came about when I first began taking an interest in consciousness through reading cognitive psychology and philosophy of mind. Blackmore and Troscianko's "Consciousness: An introduction" was especially helpful. Although I was at the time more interested in the objectivity of matters on consciousness, I was inspired to take a more subjectivist approach. At this point I adopted the viewpoint of a psychological analysis. Whether we like it or not, the hard problem of consciousness is at the heart of consciousness studies, and it is a topic which engenders antagonistic perspectives. Attacking the problem, to my mind, meant to develop a psychological theory of consciousness studies. The project began laboriously, with various written dispositions, attempting to formulate the main idea in as clear and concise a manner as possible. This was very useful when it came time to write the project description, and to the rest of the project, which germinated from this initial formulation.

In direct relation to the development of the current work, I would like to personally thank my supervisor, Professor Mark C. Price. Mark, at every turn, and at all conceivable times, never failed to provide sage advice as a consequence of his knowledge of academic writing, his well reasoned thoughts on consciousness, and his curiosity. Most of all, I thank him for challenging my way of thinking, so that I could formulate my ideas in a manner that is as clear and intelligible as possible, even to myself. Without Mark, my thesis would no doubt be as opaque as the field of consciousness studies is to the uninitiated. In his feedback, Mark went well beyond the responsibilities of a supervisor and took a personal interest in my work. Between us grew a correspondence which strengthened my own interest in, and application to, the thesis. Mark provided in-depth advice and revisions which enriched the work below, but always ensured that it remained my own.

Table of contents

Introduction

Miraculous as it may seem, we human beings are alive here in the world. We wander around and gaze out at it. Further, we appear to be able to recognize that this is the case. Not only that, we have also been given the ability to ask the most fundamental question of "why". We ask ourselves and those around us, "why am I here?", "what is *this*?", this thing that it is to be alive. More specifically, how is it possible that there is "something it is like" to *be* me, *for* me (Nagel, 1974)?[1] This deepest personal question is the question of consciousness, and whatever our ultimate motives (personal, scientific, or spiritual), we would desperately like to know the answer.

Despite our strong wish to unravel this mystery, we all imagine that we have consciousness, and are thus closely acquainted with it. However, if you ask people on the street "what is consciousness?" you will get a myriad of responses. It is to "know that you are alive", it is that you "remember who you are", it is to "understand your senses", and so on. The very concept itself seems confounded. Our philosophical understanding of consciousness is no less confused. Still, philosophy has rather neatly defined the general area of interest. Historically, consciousness has been equated with the *thinking soul* (Dolan, 2007). In his time, the philosopher Descartes coupled mental activity with being itself, and seemingly established it as an indisputable fact. We all recognize his famous *cogito ergo sum*. Though, this involved describing the soul as a detached immaterial recipient of sensory stimulation, making it a compromise with religious trends at the time (Facco et al., 2017).

Detached from such trends, in our modern understanding we say that consciousness is our "experience". This basic definition is such that it is meant to cover words like *seeing* and

[1] A glossary of some of the more technical concepts used in the paper can be found in Appendix A.

feeling at the same time—that is, the common denominator between the *experience* of feeling and the *experience* of seeing (Chalmers, 2020). Consciousness being experience is a circular definition but it also seems to be the only uncontroversial way of stating the subject matter. In fact, the definition is still very much under construction; so much so that it has encouraged opinions that a proper definition of what we are talking about must come *after* we are finished explaining the phenomenon (Sattin et al., 2021). However, despite many philosophical and theoretical disagreements it is at least minimally controversial to call consciousness "subjective first-person experience", whatever we take that to ultimately mean. The essential aspect of this definition is that it is a type of *inwardness*. This is exemplified to us in that the concept of consciousness likely used to be closer to the idea of *conscience* (Klempe, 2020). In fact, in some languages the two words are still morphologically identical (e.g., García-Castro, 2019). Still, despite these indicators, we are falling short of a satisfying definition.

The difficulties we are having in even defining the subject matter is reflected in the vast literature on consciousness. Despite the attempts of some of our greatest minds throughout distant and recent history, we still appear to be at a standstill as to what it is. The problem "seem[s] to have been around forever, yet neither science nor philosophy has been able to provide an answer" (Lamme, 2010, p. 204). Consciousness has been called "the major unsolved problem in biology" (Koch, 2004, p. xiii), as well as science at large. Indeed, some think that solving the problem of consciousness will somehow bring us considerable benefits, ushering in a sort of "new age" of science (Rosseinsky, 2019). Needless to say, we appear to want to *explain* consciousness. But what does that mean? If we cannot even define it without already trying to explain it, how do we know what to explain?

One reason why it is so difficult to conceive of an explanation of consciousness could be because our very explanations are derived from it. That is, if our subjective viewpoint (consciousness) is the basis from which we derive explanations, how can we explain that very

viewpoint (Kant, 1781/2005)? We then require a perspective-independent way of securing an explanation. Luckily, evidence conceived as independent objective stuff is the mainstay of science. Most people agree that our best current bet to explain consciousness is scientific explanation, a sentiment that motivated the emergence of *consciousness studies*. Consciousness studies is a multidisciplinary scientific field which attempts to explain consciousness using evidence mainly from cognitive neuroscience and psychology, but also includes fields such as philosophy (Francken et al., 2022). As no serious researcher in consciousness studies outright denies that consciousness is strongly associated with the brain, this evidence usually involves *physical* descriptions of neurons or *functional* descriptions of neurons. Thus, a scientific explanation of consciousness is to explain our subjective first-person experience in terms of processes which are assumed to be implemented by the brain.

The hard problem

However, consciousness seems to be a mental thing. When we think about consciousness in terms of subjective first-person experience the question arises as to why physical or functional processes should be accompanied by experience, as opposed to no experience at all (Chalmers, 1995, 1996). There appears to be a "gap" between the purely physical stuff of the world, and the purely mental stuff of experiencing that world (Levine, 1983). In other words, a key ingredient seems to be missing which allows us to go from our physical explanations of brains, to the existence of first-person viewpoints. These concerns have intermittently been called the *hard problem of consciousness* and the *explanatory gap*. It has become standard practice in consciousness studies to contrast "easy problems" with "hard problems" in that easy problems are in principle solvable by existing scientific methods, whereas hard problems appear to require something more (Chalmers, 1995). The problem of consciousness is seen by many to be such a hard problem. As they essentially constitute the

same issue, one way of combining the hard problem and explanatory gap is to say that "The Explanatory Gap illustrates … why the Hard Problem is so hard" (Revonsuo, 2010, p. 40). I refer collectively to the hard problem of consciousness, the explanatory gap, as well as the mind-body problem by the abbreviation "**HP**" (Hard Problem). In other words, I refer to the well-established academic interest in the problem that consciousness seems to be somehow fundamentally distinct from the physical world which it observes. For given that consciousness is mental and the brain is physical, how could consciousness arise from the brain? It is often asked how *physical states* could give rise to *phenomenal states* (Tye, 1999).

Theories of consciousness (henceforth **ToCs**) are employed to answer this question. It is important to mention here that when I refer to ToCs, I refer generally to *neurocognitive theories*: Theories in cognitive neuroscience that seek to explain consciousness in terms of brain processes. The reason why I do this is because neurocognitive theories are the most abundant and popular theories in the field (Sattin et al., 2021; Seth & Bayne, 2022). It is also quite rare to hold a neuroscientific theory of consciousness which is also not in some way cognitive. Such theories do exist, as well as theories which are not restricted to the biological level (e.g., quantum physics theories, or electromagnetic field theories), or even the physical level (e.g., idealist theories, or philosophical higher-order thought theories), but they are usually outliers in the theoretical landscape.

Besides pointing out the obvious correlations between brain activity and mental activity, we require ToCs to tell us the *specific* manner in which physical matter gives rise to consciousness. Many such theories have been proposed, but no theory has been widely accepted (Yaron et al., 2021). In fact, they appear to be proliferating (Seth & Bayne, 2022). It has been noted that this proliferation may be due to, for example, lack of conceptual clarity (Rosenthal, 2021) or lack of stringent criteria for theories (Doerig et al., 2021a). Alternatively, it may be because attempts to bridge the gap between physical and phenomenal states are not

intuitively understood or understandable (Price, 1996), or are otherwise still explanatorily trivial at this early stage in the field. In other words, we are not getting to what we want to know (Blackmore & Troscianko, 2018). It makes it so that the first thing we want to do when delving into the philosophical or empirical literature on consciousness is to have our *own* theory.

Although the difficulties in solving HP might spur a specialist field such as consciousness studies to launch a multitude of ToCs, to other researchers in psychology and cognitive neuroscience such issues must appear almost purely philosophical. Why should we care about theories of consciousness? The simple answer, and no doubt the motivation of many in the field, is that without a ToC which successfully diffuses or solves HP, all we have are brute correlations between the physical and the mental. For example, neural activation in the amygdala *is* correlated with experiencing negative emotion, but we currently have no idea *why*. By analogy, Newton developed the law of gravity long before Einstein ever offered an explanation to *why* nature behaves this way (Schurger & Graziano, 2022). Without an answer to issues like HP, the basis of the sciences of psychology and cognitive neuroscience is incomplete. A successful ToC is the foundational thought of psychology and cognitive neuroscience. Without it, psychology might be a bunch of random linguistic constructs with no grounding in physical reality, and cognitive neuroscience the study of mere biological matter with no reference to our lived lives whatsoever. In both these branches of science we *believe* in the relation between mind and matter. Now we must *prove* this relation to ground our beliefs.

To make this point even clearer we can consider one neurobiological theory of consciousness: Recurrent Processing Theory. In Recurrent Processing Theory, consciousness is thought to arise from recurrent activity in sensory areas (Lamme, 2010). Recurrent activity is brain activity which is highly interconnected, featuring both feedforward and feedback

connections. For instance, we can observe a feedforward "sweep" of processing from "lower" to "higher" cortical areas, for example, from V1 towards the prefrontal cortex. At the same time, feedbackward processes move from "higher" to "lower" areas, while dynamically interacting with processing levels in the forward sweep (Wu, 2018). In this sense, neural processing "recurs", a phenomenon which is thought to be necessary and sufficient for consciousness (Lamme, 2010). The question is now: Why should *this* or *any* constellation of neurobiological organization lead to a consciousness experience, as opposed to no conscious experience at all? What is it *about* recurrent activity, exactly, that gives rise to consciousness? It is easy to imagine a hypothetical world which is populated by individuals completely devoid of conscious experience, who still retain all of the complex neural organization we hear about in different ToCs (Chalmers, 1996). We are left with the conundrum that all physical explanations of consciousness appear to work perfectly fine in absence of the very phenomenon they are supposed to explain.

Disagreements on the hard problem

However, there is another way of approaching the issues we are experiencing in consciousness studies. While it has been a widely accepted convention to equate an explanation of consciousness to bridging or solving HP, there are still those who wish to construct theories while completely leaving out such a contribution (Frankish, 2016). This stance has matured under the name *illusionism*. These thinkers envision consciousness as a sort of "mere subjective experience", leaving out (and discrediting) the purely phenomenal aspect. Indeed, to varying degrees, they consider this aspect to be an illusion. To the more conventional camp, the illusionists' mere subjective experience is no experience at all, and an explanation of consciousness must include the phenomenal aspect. Though it goes under several names, I prefer to call this stance *phenomenal realism* (e.g., Van Gulick, 1994). The

opposing views of these two major camps amounts to a serious disagreement in the field as to what a ToC should explain. The issue seems to be this: We cannot agree on whether a ToC should clarify its purely *third-person* physical and functional aspects, or also its exclusively *first-person* phenomenal aspects. The argument may be seen to follow along these lines:

Premise 1: Science is an exclusively third-person endeavor

Premise 2: Consciousness is an exclusively first-person phenomenon

Premise 3: First-person and third-person endeavors cannot be united

Premise 4: Science is unitary

Conclusion: Therefore, there cannot be a science of consciousness

The phenomenal realist denies Premise 1 and imagines a science which makes room for first-person phenomena, while the illusionist denies Premise 2, and imagines a consciousness which can be explained in third-person terms (Dennett, 2018). Importantly, this disagreement appears to cause a division all the way down to empirical methods and data (Northoff & Lamme, 2020; Pinto & Stein, 2021; Signorelli et al., 2021; Yaron et al., 2021). Revonsuo (2010) writes:

> This disagreement is currently the most serious dividing line that separates different theories of consciousness from each other and also colours the interpretation of empirical results on the neural correlates of consciousness. Thus, whether a neural phenomenon that has been detected to correlate with conscious perception will be interpreted as a correlate of the actual *subjective experience* involved in perception depends largely on who interprets the results and on what background theory of consciousness it is based (pp. 222–223, emphasis in original).

This means that even in an ideal scenario where an existing or future theory (out of the myriad of theories) is basically true and a majority accepts it, there would still remain the

question of whether it chooses to answer to HP. There would therefore remain the question of whether its *explanandum* (i.e., what the explanation targets) is correct. Hence, it would be reasonable to doubt—from the viewpoint of each camp—whether it really has explained consciousness.

The meta-theoretical approach

In a science of consciousness we must first agree what we want to explain, and we must define this phenomenon (Del Pin et al., 2021; Rosenthal, 2021; Schurger & Graziano, 2022). More than in any other field, this actually leads us to reconsider what an explanation is and should be (e.g., Fahrenfort & van Gaal, 2022; Fields, 2021; Signorelli et al., 2021; Signorelli et al., 2022). These are not problems to be taken lightly. While consciousness studies have produced plenty of reasons as to how and why phenomenal aspects should be included in, or removed from, the discussion, it has usually neglected to examine why these reasons vary along these two lines. "While illusionists claim that phenomenal consciousness does not exist, many philosophers of mind regard illusionism as ridiculous, stating that the existence of phenomenal consciousness cannot be reasonably doubted. The question is, why does such a radical disagreement occur?" (Niikawa, 2021, p. 1).

For this reason we may be in need of a sort of "meta-science of consciousness". Such an endeavor may exist in latent form within recently formulated concepts such as the *meta-problem of consciousness*: The problem of why we think that there is a hard problem of consciousness (Chalmers, 2018; Frankish, 2019). This is because it puts the spotlight on us researchers as opposed to supposedly objective issues on which we disagree. However, the meta-problem has not yet been dissociated from the major camps. That is, even though the meta-problem represents an attempt to reach common ground by posing the question of "why we *think* that" as opposed to "why it *is* that", the explanations of this "thinking" is still

colored by the tacit views held by the opposing camps. As such, the meta-problem is still nested *within* each camp, and not *between* them (Sękowski & Rorot, 2022; White, 2021). I propose that beginning to dissociate this push from the major camps can be attempted by posing what I call the *meta-explanatory gap* or *meta-gap*: The problem of how to explain (and bridge) the gap between opposing positions on HP. This problem requires nothing short of a psychology or sociology of consciousness studies, for no explanation or bridge could be constructed without accommodating the characteristics of *both* camps, while relying on neither.

Agreeing on what we want to know in consciousness studies may not be as easy as simply convincing those who disagree with us; that is, the illusionist convincing the phenomenal realist, or vice versa. As researchers, it is integral that we begin with a solid foundation and build from there. I argue that the problems we experience in attempting to build this foundation lie deeper than we originally thought, namely in our own psychology. The importance of discussing HP is not to determine which stance on it is the normatively correct one. It is also not to explain why we feel or think HP is important (or not) from our preferred philosophical trench. It is not even to explain what it is or how it ultimately emerges. Since it is now affecting our science, it is first and foremost to describe how and why we human researchers systematically disagree on it. I hold that we must use psychology to remedy consciousness studies, to save the basis of both psychology and neuroscience. The end result will hopefully tell us something about why opposing theories and theorists behave as they do, so that we can inch closer to a commonly held and sorely needed theory of consciousness. “We must not ignore the psychology of the hard problem” (Price, 1996, p. 311).

This thesis represents my posing of, and contribution to, the meta-gap. Hence, it is an attempt to explain the unwanted dichotomy that we are facing in consciousness studies. The

approach I am taking here is meta-theoretical. My main line of argumentation goes as follows: We researchers have individual differences which probably strongly affect our relationship to explanations. These individual differences are explicable in terms of deeply differing dispositions towards two crucial explanatory foci in arguments. One focal point is *external*, another is *internal*. In other words, we have a *cognitive style* (e.g., Kozhevnikov, 2007) which is to begin at different ontological starting points within explanations: Some people tend to begin from physical reality and move towards subjective experience (e.g., from "photons" to "redness"), while others tend to begin from subjective experience and move towards physical reality (e.g., from "redness" to "photons"). Usually this would not be an issue, as both parties would be able to superficially agree on most things. For example, despite differences in starting points they would be able to agree on the existence of objects like rocks, chairs, and coffee cups.

However, this disposition leads to deeply conflicting stances on issues such as HP. Namely, that HP either *must be*, or *cannot be*, rationally rejected. I will make the case that this disagreement is made possible since HP is formulated as an *epistemological* issue of begging an explanation (e.g., "one just cannot see how consciousness can be physical"), and not as a potentially indisputable *ontological* argument (e.g., "consciousness cannot be physical because X, Y, Z"). In principle, some people can "see how", and some cannot. The disagreement is therefore not rational at all. This leads to different ideas about what consciousness itself is. Researchers with an *external explanatory focus* default to the position of illusionism, arguing from a third-person perspective that consciousness is reducible to physical or functional processes. By contrast, researchers with an *internal explanatory focus* default to the position of phenomenal realism, arguing from a first-person perspective that consciousness is irreducible. They both talk about "consciousness" but do not agree on the definition.

Disagreements on definitions lead to differing theories, notably, theories of consciousness. The reason why differing ToCs are problematic is because consciousness is not an *observable*, but a presupposition of observation. Consciousness as a phenomenon is not only partially, but fully *theory-laden* (see e.g., Okasha, 2016, pp. 81–82). Merely observing it depends on the theory we adopt about it. We cannot say "there it is!" without further explaining our theoretical standpoint of what that means. By analogy, the sun rising in the east, and setting in the west, is an observable empirical phenomenon. However, to a person with a geocentric model of the solar system, the sunrise and sunset looks like the sun rotating around the earth. Conversely, to a person with a heliocentric model, the same phenomenon looks like the earth rotating around the sun (cf. Anscombe, 1959/2001, p. 151). This means that empirical data cannot arbitrate between any two theories of consciousness—cannot falsify incorrect theories—since what counts as data (what we see) is fully determined by the theory. When comparing different ToCs, we are literally looking at different things, sometimes very different things, "comparing apples and oranges" (Francken et al., 2022; Pinto & Stein, 2021; Rahimian, 2022). Widely different theories developing in isolation without cross-talk then constitute the final symptoms of this causal chain. This is the difficulty that consciousness studies is facing.

Formalizing the meta-theory

In this paper I am going to expand on this broad line of reasoning by using a series of literature searches to support its most important premises. Since most ToCs are neurocognitive models, the focus will be especially on neuroscience. I have divided my approach into seven research questions (henceforth **RQs**). Below I list these RQs in the order of an argument which takes us from essential characteristics of the literature on consciousness studies, to the proposition of individual differences as an explanation for the issues we are

facing in the field. This was done in an attempt to show how my meta-theoretical approach is the logical end point of these questions. In other words, my literature searches are an attempt to corroborate a series of linked hypotheses, some of which are more commonly stated in the field, and some less so.

The final RQs were: (RQ1) Is there an abundance of ToCs in the field, or only a few well developed theories? (RQ2) Is there a consensus in the field about which theory we should accept? (RQ3) Is HP considered to be important—explicitly or at least implicitly—by the neuroscientific community? (RQ4) Is there a division in the field about whether to accept or reject HP? (RQ5) In theory and in practice, is HP best explained as ontological or epistemological? (RQ6) Are epistemological positions on HP the consequence of what is considered to be admissible arguments? (RQ7) Can admissibility in the context of HP be analyzed as the personal dispositions of individual researchers?

Stated in reverse, and in plain terms, the RQs turn into the narrative I have developed above: (RQ7) We researchers have personal dispositions, (RQ6) which drive what can possibly be admitted by us in arguments, (RQ5) to which the epistemological issue of HP is especially susceptible. (RQ4) Disagreements on HP creates a division in consciousness studies, (RQ3) which is considered to be important even in neuroscience, (RQ2) which engenders a lack of consensus around theories, and (RQ1) which is connected to there being an abundance of isolated theories in the field.

The main body of this thesis is dedicated to the attempt to answer the RQs adequately. The paper is an explorative analysis of the field of consciousness studies. More formally, it could be referred to as a *meta-narrative review* (Newman & Gough, 2020), although it also has similarities to several other types of reviews, and is closer to the development of a meta-theory. In the next section I detail the specifics of my methods. I present each RQ in turn, and under each RQ I specify several literature searches, as well as their results. After going

through the RQs, I discuss how the findings relate to the rest of the literature on consciousness, while presenting the main conclusion of each RQ. In the final part of the paper I detail why my findings indicate an obstacle for investigating consciousness empirically, and how my approach suggests a future direction for consciousness studies.

Methodological approach

Web of Science (**WoS**) was used as my only bibliographic database. I chose this database since it yields diverse papers from my main fields of interest, i.e., neuroscience and philosophy. Preliminary searches were also made in PubMed, a well-known database specifying biomedical topics, including neuroscience. These preliminary searches are not reported. During the searches, the two databases showed a substantial amount of overlap in papers. This is another reason why I limited my searches to WoS. The Google Scholar search engine was also used, albeit more as a tool to look for popular papers and citation numbers.

Further, I performed two different types of searches, one which I will refer to as *quantitative* and one which I refer to as *qualitative*. The quantitative searches look at, and compare, the number of papers that are returned from a search in order to evaluate how often certain ideas emerge in the literature. The qualitative searches go into the contents of the papers themselves to evaluate these ideas more closely. Papers' contents were rated according to a set of criteria corresponding to the RQ (e.g., on whether or not a consensus on ToCs is present in a selection of papers). Additionally, searches in Google Search and introductory textbooks were at times used as initial sources. When any of the searches were qualitative they followed the logic of PRISMA flow diagrams (per Moher et al., 2015), meaning they identified a range of papers, which was then reduced by systematic exclusions. These exclusions are reported in the text below.

Throughout exploring the RQs, additionally gathered literature (especially through citation chasing) and logical proposals are used to strengthen the RQs. Additionally, since literature searches were unsuitable for some RQs, these two latter methods sometimes form the bulk of the argument (see RQ6 and RQ7). In addition to what is reported in the main text, full versions of search strings can be found in Appendix A. Quotations used for select RQs can be found in Appendix B.

I will now go through the RQs in turn while describing my search strategies for each question in more detail. For the purposes of clarity I describe the *method rationale* for each RQ, then the *approaches*, and under each approach the *findings* of the search. The method rationale describes the general attitude which was taken toward each RQ, the approaches go into the specifics of the method, and the findings report the results of the searches.

RQ1: Is there an abundance of ToCs in the field, or only a few well developed theories?

Method rationale

For the first RQ I wanted to find good overviews of different ToCs which are regarded as relatively prominent. My strategy was first to look for representative textbooks on consciousness that contained a list or chart of different theories of consciousness, and to extract the number of theories that were listed. My second approach was to look for central and recent papers discussing aspects of a number of ToCs. That is, I looked for papers with interests *across* theories and not *within* theories. I call these *ToC-interested papers*. The next step was to extract the number of theories listed in those papers. My attempt to extract the number of ToCs was therefore based on existing overviews in the field. However, an independent search was not needed as even a preliminary search will tell us that there are quite a lot of theories.

Approach 1

For the first approach I performed a non-systematic Google Search for introductory textbooks on consciousness. After selecting three representative textbooks on consciousness that also listed ToCs, I viewed these lists and charts of theories and extracted the number of theories that were mentioned in each.

Findings. The search among textbooks revealed a relatively large number of ToCs. Blackmore and Troscianko (2018) cite Varela's (1996) two-dimensional chart showing 16 competing theories. Seager's (2016) "Theories of Consciousness" consists of 13 chapters, each devoted to one category of theories of consciousness. Each chapter goes into several specific theories as subcategories. Finally, Revonsuo's (2010) "The Science of Subjectivity" contains subchapters consisting of nine philosophical theories, and seven empirical theories, making it 16 theories in total.

Approach 2

Further, I did a semi-systematic search in Google Scholar, specifically for ToC-interested papers. To extract these articles I used the search terms "theories of consciousness" AND "models of consciousness". Since I only wanted papers that featured updated views on the field I limited the search to 2019-2022. After the first 20 papers, subsequent papers appeared to be less relevant since they started detailing specific uses for ToCs within certain fields. They therefore tended to lose the more general focus on ToCs that I was looking for. Hence, I limited the search to these first 20 papers. Throughout my attempt to answer my RQs I additionally performed non-systematic searches in WoS to affirm that this was a representative selection of papers. Within the selected 20 papers I excluded six papers (four were not interested in perspectives across ToCs, one was a re-released paper, one was not an article), leaving me with 14 papers total. The full texts of the remaining papers were accessed.

Three of the papers were rated as featuring a list of theories as opposed to only a small selection of theories, and were used to answer the RQ.

Findings. The papers identified in the Google Scholar search showed the same pattern as the textbooks, although even more theories were listed. Additionally, the papers identified in the search were much more recent. Seth and Bayne (2022) list 22 competing theories explainable in neurocognitive terms. Sattin et al. (2021) cite 29 competing theories. Signorelli et al. (2021) cite 17 non-philosophical theories.

In sum, there does appear to be a large number of ToCs in the field. This observation is confirmed by Doerig et al. (2021a) as they decry an "abundance of extremely different theories" which are "diverse in nature", and that this "contrasts with other fields of natural science, which host a smaller number of competing theories" (p. 41).

RQ2: Is there a consensus in the field about which theory we should accept?

Method rationale

RQ2 is an attempt to inquire into whether there could be an emerging consensus in the field after all, despite there being a lot of different theories. I had two approaches to this question. First, I wanted to investigate whether some theories were considered to be more popular than others, as such popularity could be taken as indicative of a degree of consensus. Second, I wanted to see if there was cross-talk between those theories, as this would indicate an attempt to reach a proper consensus. My second approach was to search the literature to see if researchers writing about ToCs considered there to be a form of consensus in the field. To do this, I looked for papers on neuroscientific ToCs that mentioned an equivalent of the concept "consensus" in their abstract.

Approach 1

Reusing the previous search comprising 14 ToC-interested papers, I selected a subset of papers which *did* feature a small selection of popular theories, and which discussed these theories. The final list contained four papers out of the 14.

Findings. The four papers discussing a small number of theories tended to select the same three to four theories. These theories were: Integrated Information Theory (**IIT**), Global Neuronal Workspace Theory (**GNWT**), Higher-Order Theory (**HOT**), sometimes with the addition of Recurrent Processing Theory (**RPT**) or Predictive Processing Theory (**PP**). It was mentioned that any such selection of theories in the field is arbitrary (Northoff & Lamme, 2020). Further, the four papers also respectively report: a diversity of theories (Northoff & Lamme, 2020), a plethora of theories (Rahimian, 2022), that it is unclear how theories relate to each other (Seth & Bayne, 2022), and that there are numerous isolated theories (Yaron et al., 2021). Rating a wide selection of papers, Yaron et al. (2021) additionally found that theories in their selection did not feature cross-talk, or even interest in other theories. In fact, it appears that individual theories rarely mention other theories (Del Pin et al., 2021).

Approach 2

Additionally, I performed a systematic search in WoS where I combined three categories of search terms: (1) theories or models of consciousness, (2) neuroscience, and (3) consensus (for the exact search string see Appendix A). Here I used Neuro* as a restrictive term in order to guide the search towards neuroscientific ToCs. Since the search is already restricted by the individual keywords detailing “consensus”, a stricter version of Neuro* was not used (e.g., Neurosci* or Neuroscience). The logic here is that I do not want to limit a search more than necessary. I also limited the search to 2019-2022 to gather recent perspectives. Twenty-three papers were identified. Six papers were excluded as they were

either not about ToCs, or only about one theory. The final list included 17 papers. The 17 papers were then rated on whether they state: (1) clear lack of consensus, (2) vague lack of consensus or the potential for an emerging consensus, or (3) consensus in the field.

Findings. The selection of 17 papers from WoS generally showed the same pattern as the first approach. Nine papers stated a clear lack of consensus in the field, eight papers stated a vague lack of consensus or the potential for an emerging consensus, and no papers stated a widely held consensus (the selection of papers and the quotations used from each abstract can be found in Appendix B). Three of the 17 papers regretted the lack of agreement on the very concept of consciousness. The other papers then varied in reporting lacking consensus on a shared neural model, the matter of reportability, physical basis, primacy of first- vs. third-person data, theory testing, and general convergence between theories.

To sum up, there appears to be a minor agreement on which ToCs are popular. However, at every junction it is mentioned that there are many such theories, and that there is no consensus on which theory to accept. In fact, it seems like ToCs disagree on most every point conceivable, even on what to explain, how to measure the construct, where and how to locate it, its necessary and sufficient qualities, and so on (Seth & Bayne, 2022; Signorelli et al., 2021). Lastly, while the selection of papers also mentions apparent agreements between theories, these usually come in the form of suggestions that will not necessarily be taken up by theorists.

RQ3: Is HP considered to be important—explicitly or at least implicitly—by the neuroscientific community?

Method rationale

What I wanted to do with this RQ was to see if the philosophical disagreement on HP was also present in the neuroscientific branch of consciousness studies, as neuroscience is the

field which contains the largest number of contemporary ToCs. My strategy was divided into three approaches.

For my first approach I wanted to see what percentage of papers discussing ToCs also mentioned HP. My preliminary searches indicated that I needed to specify a more limited selection of the literature. Therefore, I inserted proximity operators between my search terms to restrict the search (the exact search string can be found in Appendix A). Using this search as a base, I added search terms corresponding to mentions of HP and calculated the percentage of overlap between the two searches. To fully exhaust my HP target literature, I adopted several variations of HP search terms. One variation encompassed what is referred to as *phenomenal concepts* (e.g., "phenomenal consciousness" and "phenomenality"), which are closely related to HP. Two additional variations were papers that cited the two most popular papers which are unambiguously used to refer readers to HP, as citing these papers would indicate an interest in the problem.

Second, since the above searches give us a more static image of the literature, I also decided to look at how interest in HP develops over time. I therefore examined how many papers cited the two previously mentioned HP papers, that is, Chalmers' (1995) and Nagel's (1974) papers, each year from 1996–2022. Looking at increase or decrease in citations through time should give us a broader indication of interest in HP. Additionally, preliminary searches showed that almost any keyword specified in scientific databases tend to show an upward slope through time. Therefore, I attempted to statistically control for the confounding variable that scientific publications have exponentially increased over the years (e.g., Bornmann et al., 2021).

Third, as HP is a technical philosophical term, it may be that HP is not usually explicitly stated in papers (e.g., literally mentioning "hard problem of consciousness"). It could, however, be implicitly stated. Hence, to gain a more representative overview of the

field's potential interest in HP than my two quantitative approaches would provide, my third approach was more qualitative. This approach was divided into two parts. First, as they list theories' characteristics, I accessed and rated ToC-interested papers on whether they reported a ToC trying to explain a phenomenal concept, or wanting to solve HP. Second, I systematically selected a set of papers which can be considered as representative of four different ToCs, and similarly rated each paper on whether they reported the ToC trying to explain a phenomenal concept, or wanting to solve HP.

Approach 1

In WoS, I first specified a search for "theories of consciousness" OR "models of consciousness" with the restrictive term Neuro* using proximity operators. I then devised four methods of detecting the percentage of papers that were interested in HP: (1) I specified AND "explanatory AND gap" OR "hard AND problem" on top of the original search and compared this new search with the original search, (2) using the same method I specified Phenomenolog* OR Phenomenal*, (3) I looked at the amount of overlap in papers between the original search and papers that cited Chalmers (1995) in WoS (two versions found in the database), and (4) papers that cited Nagel (1974) in WoS. The WoS citations database was used here as it enabled me to compare the two searches using the build-in extraction tool.

Findings. My WoS search for Theories, Consciousness, and Neuroscience returned a modest amount of papers (154 articles). When restricted by "explanatory AND gap" OR "hard AND problem" the search specified 12 papers (12/154, 7.7%). When restricted by phenomenological concepts it specified 20 papers (20/154, 12.9%). Further, 18 of the papers cited one of two versions of Chalmers' paper (18/154, 11.6%). Lastly, 22 of the papers cited Nagel's paper (22/154, 14.2%). Thus, based on these searches, a moderate number (~7–14%)

of neuroscientific papers interested in theories of consciousness could be said to also be interested in HP.

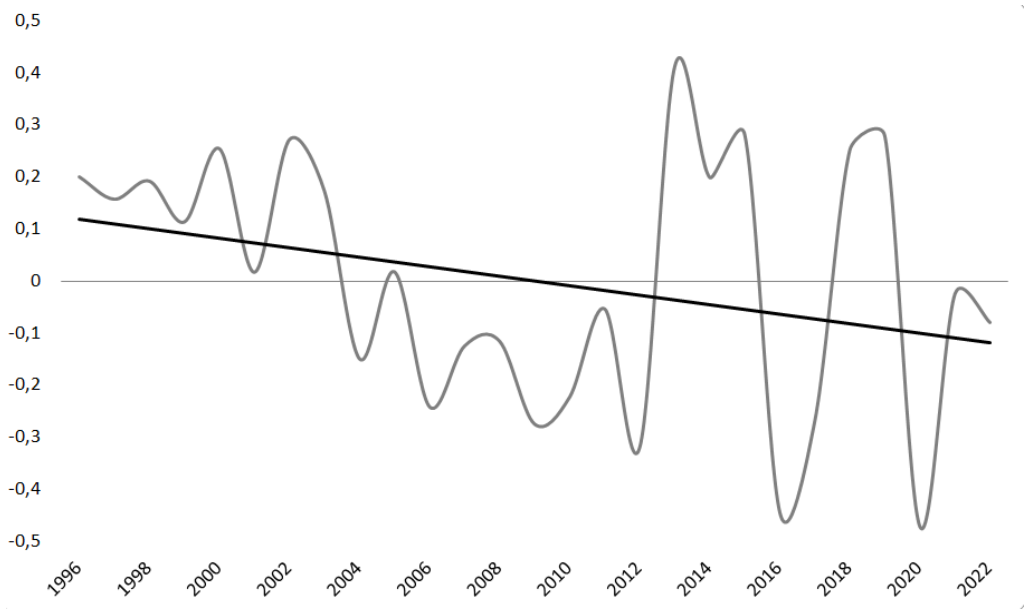
Approach 2

Following up on cross-checking with citations above, I examined trends for citations of Chalmers (1995) and Nagel (1974). Using Google Scholar, I gathered quantitative data for citations for both papers from 1996–2022 while specifying "neuroscience". Google Scholar was used in this instance, as it contained more representative data for citations. I then attempted to control for the widely known general increase in scientific publications, which could confound my results. To achieve this, I first performed three WoS searches specifying (1) only the letter "A", (2) the letter "A" in the WoS category "Neurosciences", and (3) only the term "consciousness". I then calculated and averaged Z-scores for these three scientific searches. I also calculated and averaged Z-scores for the citation trends for the two aforementioned HP papers. I then subtracted the scores of the combined "science trend curve" from the combined "HP trend curve". Finally, to see if the data showed a general increase or decrease in citations through time, I drew a line of best fit through the resulting data trend.

Findings. The trend for HP citations is shown in Figure 1. The graph shows values for the year for which the citations were gathered (X-axis), against the standard deviation of Z-scores for citation numbers (Y-axis).

Figure 1

Combined Citation Trends for HP



The data trend shows generally stable citation numbers from 1996–2012, with more variation in the data from 2012–2022. The latter range contained the years where the papers were cited the most and the least, relative to the general increase in scientific publications. The line of best fit details a small downward slope which ranges between 0.1 and -0.1 SD. Altogether, the graph shows a slightly decreasing but relatively stable trend for citations of HP.

Approach 3

My qualitative search required a more in-depth analysis of text, necessitating a smaller number of articles than a full literature search would produce. First, I reused my previous 14 ToC-interested papers. To begin with, I rated the papers on whether they explicitly reviewed ToCs or merely mentioned such theories. Six out of the 14 papers were rated as reviewing ToCs (I call these *review articles*). I then rated the six review articles on whether they (1) stated a theory as directly dealing with HP, and (2) stated a theory's main explanandum as a

phenomenal concept. Second, I again accessed all 14 ToC-interested papers, looking for mentions of specific ToCs. I then selected four theories to serve as arbitrary but representative ToCs based on their frequency of mention in these papers. When such popularity was disputable I based the selection on their frequency of being empirically tested according to Yaron et al. (2021). The theories were: IIT, GNWT, HOT, and RPT. All four selected theories were widely considered to be neurocognitive (Wu, 2018), and popular (Yeung et al., 2022). Further, I selected two introductory articles per theory based on (1) the amount of mentions of those articles across the review articles' reference lists, and (2) the recency of the article, so it would feature updated views (I call these *starter articles*). I then similarly rated the eight starter articles on whether they (1) stated the theory as directly dealing with HP, and (2) stated the theory's main explanandum as a phenomenal concept.

Findings. The qualitative search for interest in HP showed that 2/6 review papers stated a theory as directly dealing with HP, with an additional paper being unclear on the matter. Further, 4/6 review papers stated a theory's main explanandum as a phenomenal concept, again with one paper being unclear. The starter articles showed a similar pattern, with 3/8 papers stating the theory as directly dealing with HP, and 5/8 stating a phenomenal concept as being the main explanandum.

In light of these results, the general takeaway from RQ3 was that there is a modest but sustained interest in HP in the neuroscientific branch of consciousness studies. However, this interest appears to be more in phenomenal concepts rather than HP *per se*. Still, phenomenal concepts and HP are inextricably connected (Block, 1995). Additionally, some of the most popular ToCs stated the theory as directly dealing with HP or a phenomenal concept. In fact, the most popular theory in the field, IIT (Yeung et al., 2022), clearly states that it attempts to solve HP (Seth & Bayne, 2022). The results indicate that even in a hard branch of

consciousness studies like neuroscience, we do want these philosophical quandaries answered by our theories.

RQ4: Is there a division in the field about whether to accept or reject HP?

Method rationale

Following RQ3, which seeks to understand whether HP is a relevant *problem* in neuroscience as well as philosophy, I wanted to affirm or refute whether the well-known *dichotomy* in views on HP also exists in neuroscience. The first approach of this RQ was spontaneous, as it emerged during my reading of the six aforementioned review articles. The articles showed a recurring pattern, interconnecting a range of concepts which illustrate central scientific behaviors in the field. I present these groupings below in a visual manner in two *node diagrams*, one for each group of interconnections. For the second approach I performed a WoS search for neuroscientific ToCs which mentioned the two most interconnected concepts identified in the previous approach. Both concepts appearing together in the same paper would facilitate finding papers that elaborate on a potential dichotomy between them in the field.

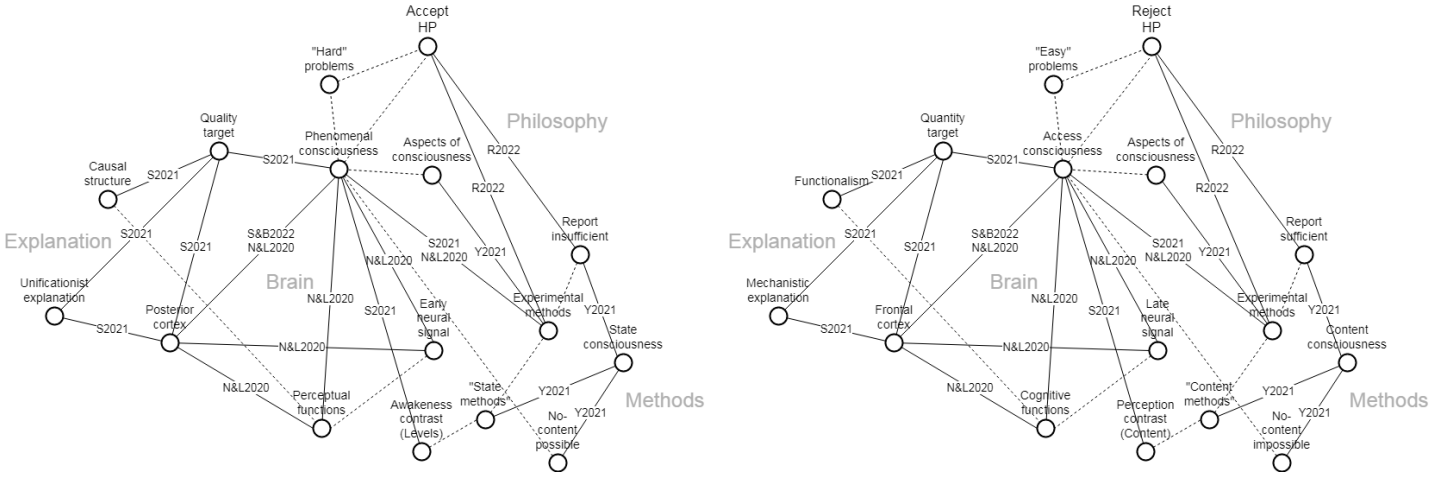
Approach 1

Using five out of my six review articles as sources, I found that concepts from these articles could be grouped around what can be referred to as two *anchor nodes*: "Access consciousness" (**A-consciousness**) and "Phenomenal consciousness" (**P-consciousness**). The two diagrams that result from this are symmetrical, as all connections are mirrored across the two anchor nodes, while no connections are implied between them. To visualize this dichotomy, I constructed the node diagrams such that each connection between any two concepts constitutes one citation from a review article, while dotted lines represent logical connections as opposed to article citations. Each diagram was roughly divisible into four

quadrants, which I labeled: *Explanation* (explanatory preferences), *Brain* (neuroscientific measurement), *Methods* (methodological approaches), and *Philosophy* (philosophical issues). The diagrams are detailed in Figure 2.

Figure 2

Node Diagrams for the Neuroscience of Consciousness



Note. The above diagram shows two sets of nodes and their interconnections. These connections are made up by abbreviated citations of papers which review ToCs. The full citations for these abbreviations are as follows: N&L2020 (Northoff & Lamme, 2020); S2021 (Signorelli et al., 2021); S&B2022 (Seth & Bayne, 2022); Y2021 (Yaron et al., 2021); R2022 (Rahimian, 2022). Further, below I list definitions of the labeled nodes used in the diagrams: **Reject HP vs. Accept HP** (the two differing philosophical positions on HP); **"Easy" problems vs. "Hard" problems** ("easy" problems are hard but solvable, "hard" problems are seemingly unsolvable); **Access consciousness vs. Phenomenal consciousness** (A-consciousness is "making information available", P-consciousness is "what it is like"); **Quantity target vs. Quality target** (Quantity targets the explanation of the contrast between a conscious and unconscious system, Quality targets the explanation of "what it is like"); **Functionalism vs. Causal structure** (Functionalism describes consciousness being generated as a function of the system, Causal structure describes consciousness being a consequence of the structure of the system in terms of causal interactions); **Mechanistic explanation vs. Unificationist explanation** (Mechanistic explanation is explanation in terms of fitting a phenomenon into a causal chain, Unificationist explanation is explanation in terms of unifying several phenomena or laws); **Frontal cortex vs. Posterior cortex** (Frontal cortex details focus on "front-of-the-brain", Posterior cortex details focus on "back-of-the-brain"); **Cognitive functions vs. Perceptual functions** (Cognitive functions are functions like attention, metacognition, and thinking, Perceptual functions are functions like seeing or hearing); **Late neural signal vs. Early neural signal** (Late neural signal is focus on measuring brain activity generally >300ms post stimulus, Early neural signal is focus on measuring brain activity generally 100-300ms post stimulus); **Perception contrast vs. Awakeness contrast** (Perception contrast is whether or not contents are perceived, Awakeness contrast is the degree to which someone is awake); **"Content methods" vs. "State methods"** ("Content methods" are scientific methods used to measure contents of consciousness, "State methods" are scientific methods used to measure states of consciousness); **No-content impossible vs. No-content possible** (No-content impossible means that consciousness must feature states with some content, No-content possible means that consciousness can feature states without content); **Content consciousness vs. State consciousness** (Content consciousness is focused on contents, such as objects, State consciousness is focused on states, such as feelings); **Report sufficient vs. Report insufficient** (Report sufficient details that reporting awareness is sufficient for consciousness, Report insufficient details that reporting awareness is insufficient for consciousness). **Experimental methods** (refers to different methods used to study consciousness); **Aspects of consciousness** (refers to focusing on different aspects of the concept of consciousness).

Findings. The two node diagrams show a matrix of connections between central concepts in the neuroscience of consciousness. These connections are complex and sometimes connect across the four quadrants. Further, they cluster around A- and P-consciousness. As previously mentioned, adopting the concept of P-consciousness is inseparable from accepting HP (Block, 1995). Conversely, rejecting HP emphasizes the concept of A-consciousness (Cohen & Dennett, 2011). It is also important to stress here that my selection of concepts was drawn from six review papers, out of which five papers are actually cited. Thus, connections between concepts may be more complex (and well supported) than what is shown in the diagrams. In a representative sample, connections could reach across the two diagrams, although whether this reflects researchers' insight or ignorance is not clear, as we may expect scientific behaviors across the diagrams to be mutually exclusive. However, the main point is well illustrated: There exists a division in concepts representing scientific behaviors in the neuroscience of consciousness, and these behaviors can be seen as a consequence of implicit and explicit positions on HP.

To exemplify this we can review a disagreement on methods. In consciousness studies, there is a disagreement on whether consciousness "overflows" conscious report (the *report insufficient vs. report sufficient* nodes), implying that *reporting* whether we are conscious either is, or is not, the best method for measuring consciousness. The disagreement stems from adopting on A- vs. P-consciousness as an explanandum. A-consciousness is more about the contrast between conscious and unconscious states, and thus lends itself to conscious report paradigms: We become conscious when we can report it. Those who adopt P-consciousness as an explanandum, however, may consider conscious report to be a confounding process which comes later than the *experience itself*, and thus interferes with locating the true neural substrate of consciousness. Proponents of this view hold that there *is* experience without conscious report, which is linked to HP stating that consciousness is

intrinsic and immediate, and not something which relies on, or is elucidated by, other functions.

Approach 2

To confirm the above division between A- and P-consciousness I performed another search in WoS. I specified a combination of the keywords: "phenomenal", "access", "consciousness", "theory", with the restrictive term Neurosci* (the exact search string can be found in Appendix A). This returned 37 papers. Eleven papers were excluded as they were not explicitly about A- and P-consciousness in relation to the field of consciousness studies. The final list contained 26 papers. To answer the RQ, I rated the papers' abstracts on whether they stated: (1) a well defined division in the field, (2) no division in the field, (3) questioned whether the supposed division between A- and P-consciousness is real, and (4) other interests.

Findings. In my WoS search, seven out of the 26 papers reported a well-defined division in the field between A- and P-consciousness views, with three papers published in 2018 and onward (7/26, 26.9%). Ten papers inquired into whether there was indeed a real distinction between A- and P-consciousness, often attempting to disprove this distinction, with six papers published in 2018 and onward (10/26, 38.4%). These two categories of papers are included in Appendix B. Another nine papers reported other interests, either elaborating on their own theory in relation to the concepts, or reporting an explanatory gap (9/26, 34.6%). No papers reported no division in the field. In addition to direct statements of a clear division in views in consciousness studies, an interesting finding in my selection of papers was that a large portion of articles doubted the reality of the distinction between A- and P-consciousness. This actually corroborates the claim that there is a division *in the field*. That is, we again find a dichotomy in views between those who accept P-consciousness as the proper explanandum

of theories (and thus accept its reality) and those who do not, or at least question its legitimacy (and thus reject or doubt its reality).

RQ5: In theory and in practice, is HP best explained as ontological or epistemological?

Method rationale

The hard problem of consciousness, or explanatory gap, is a philosophical issue. However, in the literature, it is not always clear what *kind* of philosophical issue it is considered to be. RQ5 was an attempt to find out whether HP is really *ontological* or *epistemological*. In shorthand, these are, respectively, the philosophical inquiries into what is *real* and what we can *know*. This was considered important since claims related to HP can be—and often are—taken as ontological statements, and are articulated in a manner which urges us to consider them as being indisputable. For example, HP may be formulated in such a way that we should feel obligated to accept the statement: "One is *obliged to admit* that perception and what depends upon it is inexplicable on mechanical principles" (Leibniz, 1714, as cited in Kulstad & Carlin, 2020, my emphasis); or that conscious experience is indisputable: "One might say, it makes the way pain feels into merely a *brute fact*" (Levine, 1983, p. 357, my emphasis). Hence, if HP were actually purely epistemological, this would affect the way we understand the problem. To examine whether the problem is presented by different authors in an ontological or epistemological manner, I selected and rated a diverse range of classical formulations of HP. The selection of classical formulations constituted my first approach, and was the basis of my investigation of the problem "in theory". All selected papers were seminal sources on HP, however, the oldest of these works was published in early modernity (1641), and the most recent in 1996. Therefore, to secure a selection of papers which probes contemporary literature, I also rated my previously selected starter articles (when they contained HP formulations) using the same criteria. The starter articles

were all relatively recent theoretical papers with publication dates ranging from 2010-2020. My selection of these contemporary formulations constituted my second approach, being the basis of my investigation of HP as it appears "in practice".

Approach 1

To organize my selection of HP formulations I used Chalmers' (2018) categorization of *problem intuitions*: Our intuitions that there is a hard problem of consciousness. Following Chalmers, I categorized potential sources for HP into four categories. The categories were: *Explanatory intuitions* (e.g., consciousness is hard to explain), *Metaphysical intuitions* (e.g., consciousness is non-physical), *Knowledge intuitions* (e.g., consciousness provides knowledge that is non-physical), and *Modal intuitions* (e.g., what is conceivable about consciousness). I chose this categorization structure since it makes no claims to the veridicality of problem intuitions, and therefore stays theoretically neural, making it useful to both camps on HP (Graziano et al., 2020). It also motivates a wide, and therefore representative, selection of sources. I then chose eight widely known philosophers' views on HP based on the previously listed categories. For example, one knowledge intuition is Jackson's (1982) "knowledge argument" about Mary the Color Scientist: a thought experiment wherein a person that has never experienced color (but has complete scientific knowledge about experiencing color), experiences seeing the redness of a rose for the first time (see Appendix B).

After the initial process, I performed plain Google Searches specifying the selected philosopher's last name (e.g., "Jackson"), and the philosophical concept involved (e.g., "Mary"), as well as "Encyclopedia" in order to restrict the selection to online encyclopedia entries. Viewing only peer-reviewed encyclopedias, I looked specifically for quotations which were meant to illustrate the philosophical concepts. The logic here was that quotations in

peer-reviewed encyclopedia entries should be both representative of the concepts and well-known. I limited each search to the first five result pages as the most relevant results were listed first. When encyclopedia entries did not contain quotations, I accessed HP papers which were all unambiguously considered classics within the relevant category of problem intuitions, and personally selected quotations which illustrate the concepts. From these two processes I extracted one quotation per philosopher, making it eight quotations in total (Chalmers appears twice). These quotations represented my selection of formulations of HP. To ascertain whether HP was epistemological, I then rated the quotations according to whether or not the arguments were presented in an epistemological language. Examples of quotations will be given further below (in addition to this, all quotations can be found in Appendix B).

Findings. My rating of the first selection of papers indicated that the majority of the quotations (6/8) were epistemologically oriented. Specifically, they contained a position on the insufficiency of explanation or knowledge in relation to consciousness, as well as sometimes begging an explanation of how or why physical states give rise to phenomenal states. Conversely, two quotations were not rated as epistemological. In fact, they appeared to be more ontologically oriented rather than dealing with knowledge and explanation. To illustrate this contrast I quote two papers from my selection. An example of an epistemologically oriented formulation comes from Chalmers' (1995) paper: "… even when we have *explained* the performance of all the cognitive and behavioral functions in the vicinity of experience … there may still remain a further unanswered question: *Why* is the performance of these functions accompanied by experience?" (p. 5, my emphasis). On the other hand, an ontologically oriented formulation comes from Descartes' (1641/1996) classic work:

> … there is a great difference between the mind and the body, inasmuch as the body is by its very nature always divisible, while the mind is utterly indivisible. For when I consider the mind, or myself in so far as I am merely a thinking thing, I am unable to distinguish any parts within myself; I understand myself to be something quite single and complete … By contrast, there is no corporeal or extended thing that I can think of which in my thought I cannot easily divide into parts … This one argument would be enough to show me that the mind is completely different from the body (p. 59).

The main difference in the two types of formulations seems to be between something not being *possible* or *conceivable*, which indicates the ontological orientation, and something not being *explainable* or *knowable*, which indicates the epistemological orientation.

Approach 2

Next, I again accessed my eight previous starter articles for ToCs. Reading through these papers, I extracted quotations which discussed HP. Since the results of my previous RQs suggested that HP is more implicit in contemporary ToC papers than in classical or philosophically oriented papers, I chose to look for formulations of phenomenal concepts (e.g., phenomenality) and HP author citations (e.g., of Chalmers and Nagel), rather than explicit mentions of HP. These variations of HP mentions were usually grouped together within each starter article, making selection of quotations unambiguous. Additionally, some of the papers were not focused on HP at all, and did not contain HP formulations. Seven quotations were extracted from five different papers that did contain HP formulations. I then similarly rated the quotations on whether or not they were epistemological.

Findings. My rating of the second selection of papers revealed much the same pattern as the one above. The majority of quotations (5/7) were rated as posing an epistemological

problem. Of note in this selection was that the sole epistemological concept that was used was "explanation". This may be indicative of the development of a standardization or stereotyping of HP formulations in the literature. Thus, the takeaway from RQ5 was that HP usually comes packaged in an epistemological language. We may take this to indicate that it is actually an epistemological problem which—in its initial conception and usage—has little to do with the ontological reality it sometimes purports to elucidate.

At this point it seems fair to conclude that researchers in the field of consciousness studies are having a large implicit epistemological disagreement that is causing trouble all the way down to empirical methods. This, I believe, is an uncontentious but unpopular issue. Its unpopularity may stem from the fact that it entails accepting that we are biased in science, despite our very best attempts not to be. This would constitute a natural blindspot in our investigations of subject matters such as consciousness. Science is an attempt at objectivity, so it would come as unwelcome news that there is some systematic subjectivity inherent in it. For example, although they are now widely assimilated, we only accept the writings of Kuhn (1962/2012) with a strong reluctance. The issue is also not entirely new, as similar perspectives are present in the literature on consciousness (e.g., Rahimian, 2022). The novelty of the current thesis comes from the systematic presentation of, and methodological approach to, the issue of the divisive epistemological disagreement. In addition to this, it also proposes an explanation of *how* this issue emerges, that is, an analysis of the mechanism which creates the problem, and *why* the issue emerges, that is, an analysis of the origin of the mechanism. It is to these two final questions that we turn in RQ6 and RQ7, respectively.

RQ6: Are epistemological positions on HP the consequence of what is considered to be admissible arguments?

Method rationale

The method I use in this RQ takes the form of the logical exploration of an argument, rather than literature searches, although the main point is that the logic is carried by the findings in RQ5. Namely, it is carried by the implication of a shift in our way of thinking about HP: The problem is often considered to be ontological in nature, but upon closer inspection it is heavily epistemologically laden. Further, the reasoning behind the current and following RQ is ultimately grounded in an assortment of seminal thinkers in Western philosophy, all of which have had a deep and lasting influence on science, as well as Western philosophy at large (see e.g., Dienes, 2008; Ellenberger, 1994; Klempe, 2008). These authors are often so integrated into our modern way of thinking that they are implicit in it, however, in the following paragraph I will make them explicit in the order of their contribution to the history of ideas.

I want to suggest the chronological progression of a set of highly influential ideas in Western philosophy. Importantly, the progression of these ideas runs parallel to my argument in RQ6 and RQ7 and provides a very brief philosophical background for the arguments. The progression of ideas is as follows: (1) There appears to be an overturning of ontology to epistemology; in essence, there is a "Copernican" shift from perspective-independent reality to perspective-dependency (Kant, 1781/2005). This emphasizes the active, rather than passive, role that humans play in determining what is real. (2) This perspective-dependency corresponds to two epistemologies, one *objective* and one *subjective*, which are seemingly impossible to reconcile (Kierkegaard, 1846/1994). This means that determining what is real is having a contentious perspective on it. (3) Such epistemologies are the inherent unconscious biases of us human beings (Nietzsche, 1886/2000). We can take this to mean that our

perspectives are, in a basic sense, unknown to us. (4) Specifically, these inherent biases are the consequence of individual differences (Jung, 1921/2016). This attaches the idea of an inherent contentious perspective onto the fact that people are different. (5) These differences manifest as a disagreement on whether or not the subjective epistemology can be explained, and therefore *known* publically, and not just supposedly *experienced* privately (Wittgenstein, 1953/2009). This highlights explanation rather than reality as the locus at which the division happens. (6) Such differences extend from the individual to affect science (Kuhn, 1962/2012). That is, our individual theories can strongly affect data, and therefore scientific practice.

Approach

The previous RQ implies that HP is epistemological (is about knowing), whereas it has often been considered as an ontological reality (is about what is real). The point that HP is first and foremost based on epistemology is also increasingly made by illusionists, albeit in an attempt to undermine HP arguments (Kammerer, 2016, 2021, 2022). Similarly, in the case of phenomenal realism, an ontological gap is only inferred after an epistemological gap is established (Mindt, 2017; and Park, 2013; but see Balog, 2016; and Fürst, 2011). I suspect that this point is not commonly acknowledged, as HP is often presented with the force of being ontologically real (not merely epistemologically problematic), indicating that the method by which we secured this ontology is downplayed. Indeed, this downplaying would be likely if illusionists pretend to unearth and criticize the soundness of the epistemological nature of HP as a strategy. The problem *appears* ontologically objective, but HP can be other than it appears (Stoljar, 2016).

In the context of HP, this is a shift from ontology coming first, and consequently determining epistemology, to epistemology coming first, and consequently determining ontology. “Ontology coming first” means to consider an ontological argument as primary.

While this kind of argument can also be wrong, it is axiomatic. In a basic sense, we define some initial ontological observation which drives the argument. For example, to paraphrase the type of argument proposed by phenomenal realist authors such as Descartes (1641/1996) and Nagel (1974): "It is inconceivable that experience is not real. Therefore, experience is a basic element of existence. Therefore, experience must be a brute fact. Hence, if we consider the facts, there is a problem with experience being reduced or eliminated." Since they are axiomatic, ontological arguments leave no leeway as to what constitutes an acceptable argument. In a sense, we are forced to accept an ontological argument should we accept the premises. Nonetheless, following from what was discussed previously, in the context of HP "epistemology comes first". That is, in discussions on HP, we actually consider epistemological arguments as primary. By contrast to ontological arguments, such arguments have to do with whether something is explainable / knowable. To illustrate this, we can consider an example which is relevant to the neuroscience of color vision: "There is no way to truly know the mind of another person; I cannot literally experience what their brain is supposedly experiencing. For all I know, their experience of color could be inverted from mine and I would never know. Indeed, the sum of their experience is completely private to them. Therefore, there is a separation between publicly observable things (physical objects and behaviors), and the mind, which is private" (see Byrne, 2020).

Regarding the epistemological argument above, we may now ask in Chalmers' (2018) sense: Is it the case that consciousness is hard to explain in the physical terms of neuroscience? However, note now that the answer to this question cannot follow from premises without first defining *a priori* what it means to explain something. In other words, since the epistemological argument, on its own, only references explanation / knowledge, it cannot at the same time define it. Since we cannot use the epistemological argument to define what it means to explain something, we are left with considerable leeway in formulating those

a priori assumptions. Crucially, it leaves open the question of what explanation / knowledge even is. This open question is then left vulnerable to implicit bias in the form of tacit theories (viz., "there is more to know than just physical concepts", and "there is no knowing outside of physical concepts"). In fact, Keil (1996) argued that—when starting from epistemology—making tacit assumptions about reality is unavoidable. This is very similar to Kammerer's (2022) observation that HP discussions rely on antecedently accepting contentious philosophical views that beg the question.

In sum, ontological arguments explicitly define their presuppositions, making them logically rigid, whereas epistemological arguments do not, making their acceptability subject to implicit presuppositions. In parallel to Kant (1781/2005), ontology puts a certain emphasis on a perspective-independent reality which is somehow true outside of the individual (a passive stance of having to accept). Further, since it eventually becomes necessary to justify how we as individuals gleaned any information at all from this reality, epistemology puts an emphasis on perspective-dependency *of* the individual (an active stance of determining for oneself).

As we can see in the long-lasting disagreements on HP, both in philosophy and in science, researchers actually differ on what they consider to be explainable / knowable. For example, neuroscience researchers disagree on what an explanation of consciousness is and should be (Fields, 2021; Signorelli et al., 2021). Following from the above, we can say that we differ on what we consider to be an acceptable argument. Another way of putting this is that we differ on what constitutes *admissible* arguments: which arguments we admit (accept), and which ones we do not admit (reject). Importantly, rejecting epistemological formulations of HP are also epistemological stances (Fürst, 2011; Raffman, 1995). For example, we may disagree on whether Mary the Color Scientist gains new knowledge when experiencing a red rose is an admissible argument that leads to a truth. Therefore, based on how HP is formulated

in the literature (i.e., epistemologically), accepting or rejecting HP arguably has nothing to do with what is objectively real, but everything to do with what we as readers are *willing* to accept. When we practice ontology we *must see* how, for example, consciousness can be physical. However, when we practice epistemology some people *can see* and some people *cannot see* how consciousness can be physical, and in absence of a grounding axiom there is nothing truly necessary about accepting either view (Dulany, 2014; Facco et al., 2017). This is the case in HP thought experiments such as "Leibniz' windmill" (see Appendix B). In this thought experiment one imagines that one is stepping into the workings of the mind, as if stepping into a windmill. When one observes the mechanisms inside—gears and nuts and bolts—one only ever observes a combination of parts, but never anything which can explain the mind (Kulstad & Carlin, 2020). There is the machinery of the windmill, and then there is the experience said machinery is supposed to explain, and they are fundamentally distinct. The point I have made above is that the reason why the reader might consider this fundamental distinction to hold true is not because of an objective logic, but because they *cannot see* how it could be otherwise, whereas other readers, in principle, *can*.

Findings. While often waxing ontological, HP actually puts epistemology first, making it an open question and therefore vulnerable to implicit theories. This makes variation in viewpoints likely. The conclusion I draw from this variability is that positions on HP are driven by what I call *admissibility*: The degree to which we are willing to either accept or reject certain arguments. HP (as being belief in phenomenal consciousness) is perspective-dependent; it varies with the frame of reference that we adopt about it (Lahav & Neemeh, 2022). The adoption of this frame of reference is a type of choice, and we all make this choice. The people who can see how consciousness can be physical, on one hand, do not admit HP arguments, which corresponds broadly to the stance of illusionism. The people who cannot see how consciousness can be physical, on the other hand, admit HP arguments,

corresponding broadly to the stance of phenomenal realism. We can say that people differ in their "admissibility behaviors". Some people choose option A, and some choose option B. Moreover, since they are dichotomous, moving towards one option entails moving away from the other. These are descriptions of *systematic differences* in behaviors. In light of this line of reasoning, when we observe fields in which radical and sustained disagreements rule, such as consciousness studies, we should at least subvert our expectations that researchers' theoretical choices are fully rational and detached.

RQ7: Can admissibility in the context of HP be analyzed as the personal dispositions of individual researchers?

Method rationale

As was the case between RQ5 and RQ6, RQ7 is carried by the implications of RQ6. Moreover, RQ7 takes the form of an analysis in terms of individual dispositions. I do not present the points below as the only possible analysis, but rather as establishing an initial best bet, suggestion, or heuristic towards locating the true source of disagreements on HP.

The previous RQ concluded with what I call "systematic differences in admissibility behaviors". What this means is that researchers systematically differ in their stances on HP based on which arguments they admit. Admission refers collectively to concepts such as selection, willingness, and choice. What it implies is that researchers have an analyzable bias. Surprisingly, only a small section of the literature on consciousness indicates this type of characterization. These papers occasionally mention researchers possibly being disposed towards a position on HP, or one camp on HP dismissing the other on unfounded evidence (e.g., Frankish, 2012, 2016). Other papers deal more directly with positions on HP as a consequence of a set of possible biases, for example: linguistic salience bias (Fischer &

Sytsma, 2021), theory-of-mind bias (Carruthers, 2020), and substitution bias (Miracchi, 2019).

An analysis of systematic differences in behaviors, even when conceptualized as biases, constitutes the backbone of the field of personality psychology (Corr & Matthews, 2020). Trait psychology is particularly germane. Trait psychology is the scientific analysis of individual differences in terms of our dispositions—or average tendencies—to act in a certain way (Larsen & Buss, 2017). Such tendencies are organized into dimensions and represent overt and covert behaviors. The hypothetical dataset defined in RQ6 is the same as in trait psychology. We are examining possibly inherent behaviors which vary along one dimension: Accepting versus rejecting HP. Thus, we may be able to analyze admissibility behaviors in terms of a trait, or disposition. There are three basic criteria that restrict what can constitute a trait. The behavior which is described must be: (1) consistent across contexts, (2) stable across time, and (3) individual, meaning it is not something all people share (Allport, 1931; Corr & Matthews, 2020).

Since it nestles in the cross-section between epistemology and ontology—between explanation and reality—in the following I am going to suggest the existence of what can be termed an "epistemo-ontological trait". Challenges for such a trait lie primarily in its consistency and stability, as the existence of the two camps on HP fulfill the individuality criterion for traits quite nicely. First, the challenge to consistency is that nurture could be very influential on the trait. For example, reading certain thinkers who lean heavily towards either camp could potentially tilt the trait in either direction. This would make it quite flexible and changeable across time, making it seem a bit closer to an attitude or interest. However, I suspect it is closer still to an attributional or cognitive style, as I will detail further below. Second, the challenge to stability is that the trait could also be fairly context-dependent, as it appears to be strictly limited to one philosophical problem, namely HP. However, I will argue

that it is possible to see it as a more general outlook on the world. In any case, I suggest that, within the field of consciousness studies, the time is overdue to begin exploring individual differences as a reason for radically different views on HP.

As a matter of fact, it is becoming increasingly clear that the field is teetering on the edge of such an individual differences analysis. Consider, for example, the field referred to as *the experimental philosophy of consciousness*: That is, the empirical study of philosophical ideas pertaining to consciousness. Within this field, researchers have long been concerned with problem intuitions, that is, our intuitions, usually induced through thought experiments, that there is a hard problem of consciousness (Gonnerman, 2018). Questions particularly concern whether problem intuitions are experienced by all people. Over the years this question has garnered evidence both in the affirmative (Gregory et al., 2022; Knobe & Prinz, 2008) and the negative (Díaz, 2021; Sytsma & Machery, 2009; Sytsma & Ozdemir, 2019), while others take a more agnostic view (Huebner, 2010; Wyrwa, 2022).

This discrepancy in findings, of course, may indicate that some people have problem intuitions, while others do not. This observation reemerged in consciousness studies in recent years through Chalmers' (2020) paper "Is the Hard Problem of Consciousness Universal?" in which he replies to criticisms of his (2018) paper on the meta-problem (the problem of why we think that there is a hard problem of consciousness). Among his various responses, in his reply to Irvine (2019) who suggests that "[problem] intuitions are more psychologically weighty for some" (p. 123), Chalmers acknowledges that individual differences could be a reasonable answer to the question of variability in HP positions, and thus a viable option to both phenomenal realists and illusionists.

Pertinent to this line of reasoning is the literature on the relationship between individual differences and philosophical views. Although it does not deal with problem intuitions *per se*, it has been argued that differences in personality predict systematically

divergent philosophical intuitions (Feltz & Cokely, 2009, 2012, 2016, 2019). Relevant to our discussion on problem intuitions, such findings have been found to extend to judgments about thought experiments (Holtzman, 2013). Although this view was popular early on, it is amended by later views which state that individual differences as a predictor of philosophical intuitions is probably much more complex (Byrd, 2022) and contains aspects such as *numeric interest* (Byrd & Conway, 2019) and *religiosity* (Shenhav et al., 2012). Both numeric interest and religiosity have been found to very modestly predict philosophical views (Yaden & Anderson, 2021).

In the following section I aim to nudge the field of consciousness studies over the edge towards an individual differences analysis. The first step in answering the RQ was to construct a narrative that further crystallizes my analysis of admissibility behaviors, going past the point of merely describing them. The analysis results in admissibility being comprised of three interrelated features which make up my hypothesized trait. The second step was to search three separate literatures corresponding to each of the three features. These literatures were personality psychology, the psychology of explanation, as well as social and cognitive psychology. This was done in order to see if the searches returned mentions of constructs matching each of the three interrelated features. For each feature I extracted two such psychological constructs as a way to ground my hypothesized trait in the psychological literature.

Approach

Through the discussion on admissibility behaviors I have inferred a mechanism that selects, or biases, arguments. In trait psychology, such behaviors would stem from *inherent* presuppositions. Some things just resonate more strongly with us. Why are extroverts outgoing? Well, extroversion is an analysis of people who are *inherently* outgoing, whatever

the underlying cultural or biological reasons. Further, arguments contain explanations. We all inherently begin with some presuppositions in an explanation of anything. One presupposition that could weigh quite heavily in an explanation is the basis for where the explanation should begin from. An explanation always starts somewhere, a phenomenon we can call the explanatory *starting point*. This starting point is not meant to be the very first premises or words that we use in an explanation, but rather our intuitive ontological preference. They are our deepest intuitive axioms. One reason why the explanatory starting point would heavily influence our inclinations concerning explanations is because it represents the foundation, or reasons, on which our reality rests. It is therefore possibly our grounds for beliefs and safety in the world.

I suggest that in general there are two possible explanatory starting points. Human beings tend to begin their explanations either in (1) subjective experience or in (2) physical reality, regardless of any form of proof that would confirm or deny such an intuition to them. For example, some people tend to begin from subjective experience and move towards physical reality (e.g., in the discussion of the conscious experience of color, moving from "redness" to "photons"), while others tend to begin at physical reality and move towards subjective experience (e.g., from "photons" to "redness"). With the explanation beginning in its proper place—either in subjective experience or physical reality—a transgression against this starting point in the form of a *different* starting point would feel quite unnatural. That is, when explanations transgress against our explanatory starting points, a psychological mechanism is triggered which results in us refusing to admit the arguments in these explanations. I call this hypothesized trait *internal/external explanatory focus*.

Based on the immediate narrative above, my hypothesized trait takes on the three important interrelated features: (1) explanatory starting point, (2) internal/external focus, and (3) pervasive cognitive bias. I will now attempt to substantiate the plausibility of the trait by

comparing these three features to established constructs in three respective psychological literatures. Two constructs are presented per feature. I searched one highly regarded personality psychology textbook (Larsen & Buss, 2017), one influential review article on the psychology of explanation (Keil, 2006), and several sources within social and cognitive psychology (e.g., Kunda, 1990). In the following I will use quotations from these papers to present similarities between the psychological constructs and my proposed trait.

Findings. The first feature of internal/external explanatory focus was "explanatory starting point", which I searched for in a review article on the psychology of explanation by Keil (2006). Keil references a concept called *default heuristics*, in which "… people frequently prefer one explanation to another without explicitly being able to say why. They often seem to draw on implicit explanatory understandings that are not easy to put in explicit terms" (Keil, 2006, p. 228). Additionally, for default heuristics, we appear to retreat from explanations guided by expert knowledge to implicit explanatory understandings when we impose restrictions on how much information is given to solve a problem, or the time to reflect is restricted (Kozhevnikov & Hegarty, 2001). Default heuristics appears to be quite similar to "explanatory starting points" in that it is a type of implicit explanatory inclination. More informally, there is some unconscious force that pulls us towards some explanations over others.

Next Keil refers to a concept put forward by Dennett (1987) called *explanatory stances* (alternatively, *modes of construal*).

> People may adopt a stance or mode of construal that frames an explanation. These stances are not in themselves theories; they are far too vague and nonpredictive. However, they do posit certain kinds of relations and properties, and even arguments, as central (Keil, 2006, p. 231).

Explanatory stances are similar to "explanatory starting points" in that we take certain elements for granted to frame explanations. For example, unbeknownst to us, we may adopt a teleological interpretation when framing a biological or evolutionary explanation, sometimes invoking the implicit idea of an intentional designer (Keil, 1996). Such explanations, however, are arguably better framed using the idea of a "blind designer" with no intent at all (Dawkins, 2015). Similarly, it could be the case that we begin at, for example, an external starting point in an explanation of consciousness. This might invoke the tacit idea of the physical world as the source from which consciousness is derived.

The second feature of my hypothesized trait was "internal/external focus". In their textbook on personality psychology Larsen and Buss define a concept called *attributional style*, derived from Peterson (1991), amongst others.

> Psychologists use the term attributional style to refer to tendencies some people have to frequently use certain explanations for the causes of events. … explanations for events can be either internal or external. The poor paper grade could be due to something pertaining to you (internal, such as your lack of skill) or something pertaining to the environment (external, such as the professor's being unduly harsh) (Larsen & Buss, 2017, p. 382).

Attributional style was akin to "internal/external focus" in that both concepts deal with internal/external foci for attributing explanations. It is also a type of inherent individual disposition, which is also the case for my hypothesized trait.

Larsen and Buss refer to an additional concept which appears to be close to "internal/external focus", namely *locus of control* (Lefcourt, 1991; Rotter, 1966).

> … a generalized expectancy that events are outside of one's control is called an external locus of control. An internal locus of control, on the other hand, is the generalized expectancy that reinforcing events are under one's control and that

> one is responsible for the major outcomes in life. People high on internal locus of control believe that outcomes depend mainly on their own personal efforts, whereas people who have a more external locus of control believe that outcomes largely depend on forces outside of their personal control (Larsen & Buss, 2017, p. 379).

Although locus of control emphasizes the concept of control, it shares its important core component with attributional style. That is, it is also about the tension between internal and external foci, which, again, is also the case for "internal/external focus".

The third and final feature of my hypothesized trait was "pervasive cognitive bias", for which matching concepts are present in both social and cognitive psychology. Within the field of social psychology, Kunda (1990) describes the concept of *motivated reasoning*: "There is considerable evidence that people are more likely to arrive at conclusions that they want to arrive at, but their ability to do so is constrained by their ability to construct seemingly reasonable justifications for these conclusions" (Kunda, 1990, p. 480). Motivated reasoning is similar to "pervasive cognitive bias" in that we may have underlying motives in reviewing certain explanations and in constructing new ones. This then guides our willingness to engage with certain arguments. Further, in line with my overall suggestion in the current thesis, biases such as motivated reasoning have been suggested to strongly affect scientific practice (Clark, Honeycutt & Jussim, 2022).

Finally, despite locating several constructs that aligned quite closely to respective aspects of my proposed trait, the concept that I consider to be most representative of the trait is a concept from cognitive psychology, namely *cognitive style*.

> Cognitive style historically has referred to a psychological dimension representing consistencies in an individual's manner of cognitive functioning, particularly with respect to acquiring and processing information … [That is,]

> individual differences in the way people perceive, think, solve problems, learn, and relate to others (Kozhevnikov, 2007, p. 464).

Cognitive style is related to what I have called "pervasive cognitive bias" in that it drives our individual styles of thinking and perceiving. However, as I detail in the next section, the relation between the two concepts goes beyond this single feature to represent internal/external explanatory focus as a whole.

Cognitive style is an individual differences construct. This means that people differ significantly in their cognitive processes and strategies. When it comes to similarities between internal/external explanatory focus and cognitive style, the relation between cognitive style and mental imagery as investigated by Blajenkova et al. (2006) was found to be especially interesting:

> … two types of imagers were identified: object imagers who tend to construct colourful, pictorial, and high-resolution images of individual objects, and spatial imagers who tend to use imagery to schematically represent spatial relations among objects and to perform complex spatial transformations. … object imagers encoded and processed images holistically, as a single perceptual unit, whereas spatial imagers generated and processed images analytically, part by part (p. 240).

Somewhat simplified, what these types of findings tell us is that individual differences in cognitive strategies may entail the adoption of a "holistic" versus a "reductive" mental image. Moreover, "scientists and engineers tended to be spatial imagers and … visual artists tended to be object imagers" (Blajenkova et al., 2006, p. 240). This indicates that individual differences in cognitive style can drive our interests and occupations, in the sense that the cognitive style in question is matched with the corresponding interest or occupation. Researchers have taken a particular interest in applying questions of individual differences

(including cognitive styles) to who is likely to become a scientist (Simonton, 2009). Such interests are also extended to areas dealing with mental imagery in cases outside of the normal spectrum, namely to people with aphantasia (the absence of visual imagery) and hyperphantasia (the abundance of visual imagery). In particular, people with aphantasia appear to choose scientific / mathematical occupations, whereas people with hyperphantasia seem to more frequently pursue creative professions (Zeman et al., 2020). This runs parallel to findings in personality psychology, where it has been argued that personality traits can affect chosen occupations, for example, that scientists are generally lower in the Big Five trait Openness (see e.g., McCrae & John, 1992), than non-scientists (Feist, 1998).

In addition to individual differences in mental imagery driving interest *towards and away* from science, the idea has emerged that such differences can drive perspectives *within* science. Specifically, it has been proposed that individual differences, for example, in mental imagery, can implicitly influence scientific theorizing, especially in young fields when the evidence is sparse and the field is open (Reisberg et al., 2003). To illustrate the potential severity of the consequences that differences in mental imagery can have on scientific theory, Faw (2009) considers the cases of the famous psychological theoreticians John Watson and Edward Titchener. In his writings, Watson reported that he had no mental imagery whatsoever, whereas Titchener was said to have exceptionally vivid mental imagery. It is then significant that these individual differences may well have driven the very scientific paradigms in which a large portion of researchers worked for a time, namely, behaviorism and introspectionism (Faw, 2009). This has been shown to be the case for differing theoretical positions in the early days of mental imagery research (Reisberg et al., 2003). Recently, this argument has been applied to the cognitive sciences:

> Hidden phenomenal differences may also help us to understand the behavior of cognitive scientists. Assuming that others are just like us means that a

> researcher with vivid visual imagery is likely to place imagery in a more theoretically central position than someone with poor imagery (Lupyan et al., 2023, p. 5).

This discussion is particularly germane to consciousness studies. Could admissibility behaviors for HP—which we can take to drive theoretical positions in consciousness research—be a consequence of individual differences in cognitive style? The point I have attempted to make in the current RQ is "yes", and such cognitive styles may manifest in the form of different explanatory starting points.

However, here I wish to distinguish my own proposal from the research on mental imagery in one important aspect. I am not making the claim that some researchers are incapable of understanding or accepting phenomenal consciousness because they are somehow cognitively deprived, nor am I stating that researchers who do accept phenomenal consciousness are somehow abnormally sensitive. At a minimum, I am proposing individual differences in an *explanatory* component which plays out in arguments, not in an experiential component which puts an obvious limitation on our mental lives. Antagonistic diagnoses may emerge between the two camps on HP, but they are not very helpful (Niikawa, 2021). Indeed, before the onset of comprehensive individual differences theories, for example, that of extroversion/introversion, opposing dispositions must have seemed impenetrable to each other. To be sure, such theories were initially developed to engender mutual intelligibility, and not to classify individuals. As the famous Carl Jung puts it in an interview by Segaller (1957/2000): "The classification of individuals means nothing at all. It is only the *instrumentarium* for the practical psychologies, to explain, for instance, the husband to a wife, or vice versa". Similarly, I have argued above that the instruments we require to facilitate mutual understanding between the two camps within consciousness science lie latent in individual differences psychology.

In conclusion, there does appear to be similarities between internal/external explanatory focus and well-established constructs in psychology. In addition, my hypothesized trait appears to be well-suited to be put in terms of a cognitive style. Based on these observations, I conclude that admissibility behaviors in consciousness science *are* analyzable in terms of an individual differences construct. The construct further reasonably fulfills the three criteria for conceptualization as a trait: consistency, stability, and individuality. Internal/external explanatory focus is a hypothesized trait which is *stable* across time in the sense that nurture may heavily affect the trait while still resembling more of a cognitive style and not a personality trait *per se* (see Riding, 1997). Historically, however, the main proponents of different camps on HP tend to hold firm on their individual positions (e.g., Dennett, 2017, p. 2). Further, the trait is *consistent* across contexts in the sense that it could speak to our biases in explanations as such, and thus transcends the context in which it is initially discovered, namely HP. Lastly, it is *individual* in that the extremes of the dichotomous camps represent the extremes of the continuum which individuals may inhabit.

Discussion of findings

In this section I will first summarize the findings of the RQs, and then attempt to relate my findings to the broader literature. To take a more extensive view of the thesis, the RQs have been grouped into pairs (except RQ7) so that each pair represents one of four main aspects of the paper. The aspects were: theories of consciousness, the hard problem, epistemology, and individual differences.

RQ1 and RQ2: Theories of consciousness

The first and second RQs demonstrated that the literature on consciousness contains a proliferation of widely differing theories on which there is little to no consensus. Most of my review articles explicitly drew attention to the issue of heterogeneous perspectives among

theories (e.g., Northoff & Lamme, 2020; Sattin et al., 2021; Seth & Bayne, 2022). This has been the trend all the way up to the most recent literature (e.g., Fesce, 2023). Not only are there wide heterogeneities between theories, they also disagree on specific issues, such as their explanatory target (Yaron et al., 2021). As Revonsuo (2010) notes: "… theories are so diverse that it is not even clear whether they all talk about the same thing when they use the word 'consciousness'" (p. 176). Regardless, many theories appear to target phenomenal consciousness, and seem to believe that they have bridged HP in some capacity (Northoff & Lamme, 2020; Signorelli et al., 2021). However, the observation has been made that such theories may feature intentionally vague assertions in an attempt to camouflage an ignorance on philosophical problems (Doerig et al., 2021a). Such avoidant behaviors could be a root cause for some researchers later decrying that ToCs simply miss their own explanatory target (e.g., Schurger & Graziano, 2022; Seth & Bayne, 2022).

Although diversity in science is sometimes good, we obviously seek to hone in on one true theory, to the degree that this is possible. There are not twenty widely differing theories of evolution, but one widely accepted one. Not all theories of consciousness can be true (Doerig et al., 2021a). This means that researchers in consciousness studies should be interested in ameliorating the issue of lack of convergence. Some suggest that lack of convergence is due to confirmation bias in confirmatory experiments, that is, each theory largely confirming its own observations (Yaron et al., 2021). In this vein, ToCs tend to develop in isolation without cross-talk, meaning it is very difficult to arbitrate between them (Del Pin et al., 2021; Seth & Bayne, 2022; Yaron et al., 2021). Apparently, this has also been the case historically (Michel, 2019). Further, since there is little cross-talk between theories, one theory succeeding does not mean the degradation of others, as is usually the case in more mature fields (Signorelli et al., 2021; Yaron et al., 2021). Building on these observations, others suggest that lack of convergence is due to the lack of work on comparing theories (Del

Pin et al., 2021), or to there being a division in theories' core assumptions (Signorelli et al., 2021). These considerations are what sets the stage for the primary issues in consciousness studies. In other words, the last thing we need in consciousness science is *another* theory of consciousness (Seth & Hohwy, 2021; Wiese, 2020; Yurchenko, 2022).

RQ3 and RQ4: The hard problem

Following from the third and fourth RQs we can say that there is a sustained division on HP even in neuroscience. Famously, there has been a long-standing and strong dichotomy on HP in philosophy (Chalmers, 1996; Dennett, 2018). The division we find in neuroscience can be seen as a more moderate version of this dichotomy. While not dealing with philosophical issues *per se*, it plays out in different theories being derived from different explanatory targets. Some theories follow from A-consciousness, while others derive from P-consciousness (Lamme, 2018; Promet & Bachmann, 2022). As previously discussed, both of these explanatory targets are symptomatic of opposing positions on HP. As each camp argues for the integrity of their own explanatory target, the field reaches a stalemate on whether A- or P-consciousness is the correct explanandum for a ToC (Lamme, 2018; Michel, 2019). Attempts have been made to move this stalemate forward by taking different approaches, such as posing the meta-problem, but little agreement has yet emerged between the two camps (Sękowski & Rorot, 2022). Importantly, subscribing to different explanatory targets appears to lead to the adoption of widely different methods and concepts, creating a fragmented image of the science (Rahimian, 2022; Yaron et al., 2021). One example of this division is the disagreement on which general brain area is necessary and sufficient for consciousness: for example, the (pre)frontal cortex (Odegaard et al., 2017), or the posterior and temporal cortex (Boly et al., 2017). Following this general line of reasoning it has been argued that positions on HP is what drives theoretical positions on consciousness (Sękowski & Rorot, 2022). In

turn, these theoretical positions are thought to affect how theories define and interpret data (Pinto & Stein, 2021).

Moreover, it has become a point of contention whether positions on HP result from prior exposure to philosophical arguments (e.g., Frankish, 2012; Fischer & Sytsma, 2021), or result from widely held lay assumptions that cannot be reasonably doubted (e.g., Graziano et al., 2020; Pinker, 2007). As HP positions may well drive theoretical stances, the contention develops into the question of whether or not we ought to trust our assumptions about HP. Critics of HP then question whether bringing researchers into an understanding of the hard problem systematically hinders us from being able to develop ToCs which are empirically adequate (e.g., Lau & Michel, 2019). However, regardless of whether HP is true, it is a widely held position in consciousness science. Some surveys tell us that over 60% of philosophers and consciousness researchers believe that there is a hard problem of consciousness (Bourget & Chalmers, 2021; Francken et al., 2022). In light of these perspectives, the literature on consciousness reveals a serious disagreement on the fundamental question of what consciousness is to begin with, and therefore what types of methods, measurement, and explanations should be used when investigating consciousness scientifically. To individual neuroscience researchers—with no overarching criteria to guide them—selecting a theory to work under can be likened to choosing and sticking with a football team.

RQ5 and RQ6: Epistemology

The fifth and sixth RQs inquired into the nature of HP, concluding that HP is an epistemological matter which is largely determined by our willingness to accept central arguments, and not by external rational methods. I called this phenomenon "admissibility behaviors". The concept of willingness that supplants rationality implies that disagreements on HP are about biasing towards and away from epistemological HP arguments. We adopt

certain frames of reference when dealing with HP (Lahav & Neemeh, 2022). Additionally, we may not be conscious of the fact that we adopt these frames of reference (Wyrwa, 2022). A more descriptive term for this phenomenon is *tacit theories*, our implicit assumptions or positions, which latch onto largely open questions such as HP. Some people strongly approve of the central arguments of HP, while others strongly disapprove. The relative strength of our beliefs that HP either does or does not exist could be responsible for the two camps on HP (Balog, 2016). In fact, it may cause us to reject the opposing camp's ontological standpoint *a priori* (Facco et al., 2017). Tacit theories do not do us any favors in science. As Rahimian (2022) notes, an implicit position on HP is just a poorly formulated position on HP, and these positions are unjustified assumptions which could slow down scientific progress. Thus, to begin with, the most important position a ToC can make explicit is its position on HP (Cheng et al., 2022; Pinto & Stein, 2021).

RQ7: Individual differences

The seventh RQ contained the culmination of my analysis of the field of consciousness studies. The analysis suggested the existence of an epistemo-ontological individual difference which can unconsciously drive our theories of consciousness. The individual difference takes the form of a trait-like cognitive style which biases us to start at an internal or external point of view in explanations, namely internal/external explanatory focus, which matches the two opposing positions on the hard problem of consciousness. Although their approach differs somewhat, Graziano et al.'s (2020) view aligns with this idea, suggesting that our conception of reality depends on an inherent disposition towards the source from which we draw this reality, and that this could influence our positions on HP. Similarly, Facco et al. (2017) argued the general point that historical, cultural and inherent human tendencies bias metaphysics, which biases one's axioms, which in turn biases

questions ultimately affecting science, such as HP. Although the picture that I have derived from answering the RQs is simplified, and both the representative literature and the answers to the questions are probably more complex, my approach yields the central and oft-avoided insight that we are not necessarily "the masters of our own house" in scientific theory and practice. I now turn to the implications that these findings have for consciousness science.

Implications of findings

First off, my findings have implications for the way that we perform the empirical science of consciousness. Though at this point, as scientists, we should be shaking our heads. Normally, empirical science tends to dilute individual differences of researchers through methods of distancing researchers from their research object, and through statistical averaging. In the case of consciousness studies, how could individual differences be a problem for empirical science?

Theory-ladenness: A challenge to agreement in *a posteriori* terms

In consciousness science we are harboring strongly opposing presuppositions which lead to differing ToCs. The way that empirical science usually arbitrates between opposing theories is by continuously doing empirical measurement to corroborate and falsify said theories. Falsified theories gradually fall off the map, and corroborated theories remain. This means that scientific ToCs should be *falsifiable* (Cohen & Dennett, 2011; Hanson & Walker, 2021; Kleiner & Hoel, 2021; Seth & Hohwy, 2021). Granted that ToCs are falsifiable, one assumption seems to be that we can just "shut up and measure" (Pinto & Stein, 2021, p. 97). The rest would then be left up to the slow progress of science. From this point of view, researchers hope that we will eventually arrive at an *a posteriori* agreement on consciousness via empirical measurement (Doerig et al., 2021a; Klein, 2021). However, following the previous discussion on lack of consensus on ToCs, theories seem to be developing in

isolation, while largely confirming their own observations. Widely differing theories with opposing presuppositions are all enjoying their own isolated empirical corroboration (Yaron et al., 2021). This means that, in practice, empirical science is not doing what we want it to. It is not creating a common ground for theories to fight it out.

It seems to me that consciousness science has three available tools for countering opposing presuppositions of theories. The first is to attempt to pit theories against each other (Seth & Bayne, 2022; Doerig et al., 2021a). This would presumably mean that the best theory outcompetes the other(s). The second is to attempt to combine theories with *similar* assumptions (Graziano et al., 2020). This would potentially mean that the combined predictive power of the remaining theory could outcompete rivaling theories. The third is to attempt to combine theories with *different* assumptions (Northoff & Lamme, 2020; Safron, 2020). This would mean that different assumptions are just different perspectives on the same thing, and combining them would give us a fuller picture of the phenomenon. (We may note, however, that the divergence in assumptions emerges exactly because the opposite view is held to be wholly insufficient.) In the end all these methods will have to empirically compare one theory to another to find out if it is a better explanation to the phenomenon in question. In other words, they will have to test theories against each other. Since it is having problems with theory-testing, it has been suggested that consciousness studies should adopt the practice of *adversarial collaborations* (Cleeremans, 2022; Melloni et al., 2021). In adversarial collaborations two (or more) theories collectively define *a priori* their own theoretical predictions about a future empirical observation, and how this observation would corroborate or falsify their own theory. They then run an experiment to see which theory wins out. It has been noted that such collaborations function as a remedy for researcher bias (Clark & Tetlock, 2022; Clark, Costello, Mitchell & Tetlock, 2022). However, I will now argue that adversarial

collaborations fail as a general strategy for solving the theoretical issues that face us in consciousness science.

We may ask ourselves: Why is empirical science failing to create a common ground for arbitrating between theories in consciousness studies? One answer is to consider Kuhn's (1962/2012) concept of *theory-ladenness*. To illustrate, two scientists can point to a celestial sphere in the sky, and name it "planet" and "star", respectively, based on their own theoretical presuppositions, for example, of orbital patterns (Dienes, 2008, pp. 36–37). That is, their observation of the object is determined by the theory they adopt about it. However, common to both scientists is that they can *point* to a thing in the sky. The thing is an *observable*. In other words, it is unambiguous what we mean when we point at it. Thus, the observable has only *partial* theory-ladenness, which leads us to being able to reach an eventual agreement on what it is through theory-testing. But when it comes to consciousness, as Velmans (1996) has put it: "where does one point, when one is pointing at consciousness?" (p. 183). Nowhere. While it can be theoretically inferred, consciousness is not an observable in the sense of observing an object (Pinto & Stein, 2021; Schurger & Graziano, 2022). As opposed to the celestial sphere, we cannot point to it to agree where and what it is. Thus, the mere observation of consciousness relies on the theory of what it is. We can say that consciousness has *full* theory-ladenness. I will now expand on this observation.

Contesting the statement that consciousness is theory-laden, there have been some attempts to describe consciousness as a "clear empirical phenomenon", particularly by Doerig et al. (2021a): "We wake up every day and change from an unconscious to a conscious state. Obviously, there is something to explain. … Because of these clear empirical phenomena, there is no need to posit a theoretical definition" (p. 41). However, it appears to me that this is an equally theoretical statement, namely the position that "consciousness is a contrastive waking state". Such descriptions are as intuitive and contentious as the phenomenal realist

position of Nagel (1974) that consciousness is "what it is like" to have an experience. Statements about consciousness are not empirical facts (Haun & Tsuchiya, 2021). As I have put it previously: Consciousness is not an observable, but a presupposition of observation; we cannot say "there it is!" without further explaining our theoretical standpoint of what that means. This makes it difficult or even impossible to observe with only partial theory-ladenness, making it equally difficult or impossible to agree on a theory of consciousness in a perspective-independent way, as is likely the case in the rest of science. Since the very observation is determined by the theory, any two theories of consciousness, granted that they have different assumptions, are technically not talking about the same thing. Sometimes we have very different, even opposing assumptions, and are therefore talking about very different things. One ToC's confounding variable is another's essential observation, as is the case for examples such as: conscious report (Rahimian, 2022), attention (Del Pin et al., 2021), and activity in the visual cortex (Revonsuo, 2010). In fact, two different ToCs (e.g., GNWT and IIT) can apparently interpret the identical dataset differently (Mashour et al., 2020). One reason why adversarial collaborations would fail in consciousness science is because when comparing theories corresponding to A- and P-consciousness, respectively, we are "comparing apples and oranges". The case can be made that the two types of theories really have nothing to do with each other, even if they use the same label, "consciousness". Adversarial collaborations assume that the theories that are to be compared both investigate one and the same phenomenon, but this is necessarily *not* the case in consciousness science. By analogy, it would make little sense to empirically test social psychological theories of "attribution" and "cognitive dissonance" against each other, because one does not mutually exclude the other empirically.

In summary, the process of reaching *a posteriori* consensus in consciousness science is compromised by the inherent theory-ladenness of consciousness as a phenomenon. I hope

the above proposal has made it clear that we must actually secure an *a priori* agreement on consciousness to get anywhere in the science. It then follows that we should be fine if we define and agree on our core assumptions beforehand (Francken et al., 2022; Northoff & Lamme, 2020; Rahimian, 2022). However, following from my proposal of internal/external explanatory focus, a spontaneous and genuine agreement on a set of assumptions defining consciousness is unlikely to occur, given our deep differing predispositions toward *opposing* assumptions. This brings me to the second implication that my findings have for consciousness science.

Meta-gap: The call for agreement in *a priori* terms

The current thesis is an attempt to draw attention to the fact that we have a seemingly unsolvable dichotomy in consciousness science which plays out at multiple levels, from philosophical issues to empirical methods. I have proposed that researchers, as well as people in general, have individual differences which manifest as deep differing explanatory foci. These differing foci lead to differing philosophical presuppositions. Crucially, they engender the dispute on HP. This dispute then leads to different ideas about what consciousness is, which in turn results in fundamentally differing ToCs. Differing ToCs constitute a serious problem since consciousness as a phenomenon is fully theory-laden, meaning the theory fully determines what counts as data. This then leads to incommensurable scientific observations, possibly being the primary cause for dissent in the field.

This causal chain describes my analysis of opposing viewpoints in consciousness science. Not only is there an apparent explanatory gap between physical and phenomenal states, there is also an explanatory gap between our *positions* on this explanatory gap. I call this phenomenon the *meta-gap*. The idea behind the meta-gap is that it makes itself apparent through a sort of meta-cognition. I have formulated internal/external explanatory focus as a

disposition which is inherent in our viewpoint. In other words, it is an "attitude of consciousness". When we then begin to study consciousness as a scientific object, it follows that we turn this attitude *of* consciousness *on* consciousness. We move from a first-person perspective to a so-called third person-perspective on first-person perspective. This is consciousness attempting to observe itself. When we use our consciousness to observe itself, all we do is reflect its inherent attitude back at it, all the way down to empirical methods, causing the disagreements that we experience in the field.

For example, one neuroscience theorist may have a disposition towards an internal explanatory focus. Therefore, they tend to begin their explanations from the standpoint of subjective experience. HP arguments are about the irreducibility of experience, which justifies beginning from experience. Thus, the theorist finds them attractive and conclusive. HP arguments are admitted, and HP is seen as a real problem that must be solved. This then leads to the seemingly obvious move of adopting phenomenal consciousness as an explanandum for a theory. They would say, in the epistemological language of begging an explanation: How could consciousness be anything other than phenomenality itself? The theory that develops from the adoption of P-consciousness as an explanandum, then highlights locating the explanation and neural correlates of *experience itself*. By contrast to an A-consciousness theory, a P-consciousness theory such as Integrated Information Theory (Ruan, 2022) tries to get away from the general emphasis of frontal lobe functions related to, for example, attention, working memory, or meta-cognition, as such functions are involved in bringing us from an unconscious to a conscious state, granting us *access to experience* (Northoff & Lamme, 2020).

The P-consciousness theory is then led to emphasize other brain areas and functions. The search tends to lead us to sensory areas, for example, occipital lobe, in which visual percepts are formed. Such percepts arguably resemble more of an immediate or "raw"

experience which is conceptually associated with *experience itself*, as opposed to the later cognitive processing of said experience in frontal lobe (Northoff & Lamme, 2020). Focusing on occipital lobe leads us to adopt explanatory strategies and measurement paradigms which match the neuroanatomy of this area. In IIT, the explanatory strategy focuses on the "posterior cortical hot zone", which is thought to be especially good at integrating information, a process which IIT proponents consider necessary and sufficient for consciousness (e.g., Tononi et al., 2016). fMRI measurement then fixates on brain activity in this posterior hot zone, and due to neural activity reaching posterior zones earlier than anterior zones, EEG measurement emphasizes "early" neural spiking in the range of 100-300ms post stimulus as meaningful data (Northoff & Lamme, 2020). On the other hand, a proponent of an A-consciousness theory such as Global Neuronal Workspace Theory (Ruan, 2022) would look at the data from IIT experiments and either consider the data as confounding factors, or interpret them in the light of their ultimate influence on frontal lobe systems which they hold to be necessary and sufficient for consciousness (e.g., Mashour et al., 2020). Other concepts are emphasized, namely the "global workspace", different anatomical regions are probed, and EEG measurement focuses on later spiking in the post-300ms range (Northoff & Lamme, 2020). Measurements across the two theories turn a blind eye to what they see as irrelevant data. Whatever the rival theory is doing may be a great explanation of some other phenomenon, but *that* is not considered to be *consciousness* (Schurger & Graziano, 2022). To GNWT, "information integration" may perform a function of *perception*, but not consciousness. To IIT, "global workspace" may perform a function of *attention*, but not consciousness. We are left with incommensurable observations between the two types of theories, while they may both hold that they have, in part, explained consciousness.

When we experience other people's viewpoints, especially ones we disagree with, it tends to follow that we become self-aware of our own viewpoint. Perhaps our own

perspective is just one among many? If we then attach a label to our own perspective (or attitude of consciousness), for example, "internal/external explanatory focus", we can formulate it as a bias. We turn it into an attitude of which we are conscious, as opposed to unconscious. In summary, when we collectively turn our two *unconscious* attitudes *of* consciousness *on* consciousness, we remain fixed in our own perspective and bring about the two camps and the disagreements and dissent in the field. However, when we turn our two *conscious* attitudes *of* consciousness *on* consciousness, we break out of our fixation and become aware of the relation between our own attitude and that of others. It gives us the means to formulate the insight that we are being "subjective about subjectivity". Being subjective about subjectivity implies that our own viewpoint on subjectivity is relative to other viewpoints.

In common with the explanatory gap, the meta-gap presents within itself its possible solution in bridging this gap. Somewhat simplified, bridging the meta-gap would mean to understand and ameliorate conflicting positions on HP, that is, between phenomenal realists and illusionists. It would entail literally overcoming our individual differences *before* we make an essential observation about consciousness, not *after,* as would be the case in the head-to-head competition in the rest of science. Therefore we must construct a common ground beforehand. The creation of this ground is probably hard given consciousness inspires strong (perhaps incommensurable) convictions on all sides. As Chalmers (1996) has put it: "Perhaps our inner lives differ dramatically. Perhaps one of us is "cognitively closed" to the insights of the other. … In any case, once the dialectic reaches this point, it is a bridge that argument cannot cross" (p. 151), and therefore, "We may simply have to learn to live with this basic divide" (p. xi). I take a slightly more optimistic view. As scientific publication grows exponentially, we can no longer afford to remain divided on fundamental issues. The foundation upon which we *must* build consciousness science is a common observable, not just

a common label. While agreeing upon this common observable is hard, it is a tractable problem. For one, to avoid dividing the community yet again, this new project should proceed in a manner which is impartial to whether HP exists. By analogy, a married couple consisting of an extrovert and an introvert can reach a provisional agreement on how long to stay at a party. A marriage is all about giving and taking, and we are all wed by our involvement in science. The integral change is becoming aware that we are all different types of people, and then navigating extant problems based on this insight, and not by indiscriminately imposing our own viewpoint, even if that is what we would really like.

The meta-gap is the novel formulation of a fundamental problem. I propose that giving it some serious thought is the most essential approach one could adopt in consciousness studies. The reason for this is that our psychological profiles are reflected in our scientific work, especially in sciences where observation of the target phenomenon is difficult or impossible, as *common* experiences cannot arbitrate between *individual* experiences. When the phenomenon is invisible, questionable, or otherwise unobservable, tacit ideas born of our individual psyches are given free rein in defining this phenomenon. The chaos of theories in the field reflects the ways in which we are all different. Further, the specific disagreements in the field reflect the ways in which some of those differences vary systematically. In other words, consciousness science is about as diverse as a group of conscious individuals. We must now attempt to navigate this diversity with the tools that we have at our disposal.

Limitations

In this section I address three main methodological limitations that could be leveled at the thesis. First, the thesis, while partly describable as a review, also involves the construction of a theory, and therefore does not follow a conventional review methodology. However, there seems to be a lack of precision in defining types of reviews, as well as a great deal of

overlap in review types (Grant & Booth, 2009). Incidentally, the thesis overlaps with several such review types, such as critical-, mapping-, rapid-, and meta-narrative reviews (see e.g., Grant & Booth, 2009; Newman & Gough, 2020).

Second, since the thesis only had a single author, the process of selecting and rating quotations in the RQs could be considered somewhat subjective. A further step in strengthening these ratings would be to perform text ratings across several researchers using set criteria. The objectivity of the ratings could then be confirmed via a calculation of inter-rater reliability. Unfortunately, this was seen to be beyond the practical scope of the thesis.

Third, this thesis is a psychological analysis of a neuroscientific field which is philosophically inclined. Strictly speaking, this could be said to deviate slightly from a neuroscience master's thesis. However, neuroscience permeates the majority of the discussions that are covered, even if it is not always explicitly mentioned. The thesis, like cognitive neuroscience, is an inquiry into the issues that emerge when we confidently say that there is more than a mere correlation between the cognitive domain (consciousness), and the neural domain (physical reality).

Concluding remarks

The current paper delves into the neuroscience, philosophy, and psychology of consciousness studies. Consciousness, or subjective first-person experience, is in many ways a foundational topic. It seems to be one of the core aspects of what it is to be alive, and also permeates our knowledge structures, particularly inspiring the mind sciences. Furthermore, it also itself constitutes an object of science. With the tools of neuroscience at hand, we seek to establish the connection between subjective first-person experience and physical reality. Since we already build a lot of our knowledge on the assumption of this connection, we seek to

establish a theory which explains how to go from the former to the latter, or vice versa. However, we are having problems in agreeing upon such a theory.

My strategy in formulating this point has been to present the main problems of consciousness studies, and then to formalize and corroborate a set of theoretical stepping stones which allows us to go from problematic characteristics of the field, to individual differences as a reason for these issues. Although it may seem like the most obvious fact to all of us, observing consciousness as a phenomenon engenders widely disparate viewpoints. In science, the call for agreement is often uttered but rarely heard. This may simply be because we have a strong belief that our own perspective on things is the right one. Nobody ever made any headway without being headstrong. However, in a community this type of disregard can also be disruptive. What could help this situation is to become a bit more self-conscious. This paper attempts to help us do just that.

People disagree about consciousness enough for there to be dozens of antagonistic neuroscientific theories of consciousness. We should take this as an indication that something is going on, and not that we need another theory. We are all individuals, and that means that we all have individual psyches. When we charge headlong into an observation of consciousness, we tend to assume that everybody see things the way we do. Our inherent presuppositions can then become highly problematic. However, with just enough self-consciousness to counteract the instinctive reaction to conclude that everybody is like ourselves, we can see that we were unwise in our headlong charge. Other people see consciousness differently. As if looking into a mirror, we turn our observation of consciousness into an *observation of our observation of consciousness*, and become self-conscious. To my mind, this is the first step in reaching an agreement on consciousness.

References

Allport, G. W. (1931). What is a trait of personality? *The Journal of Abnormal and Social Psychology*, *25*(4), 368–372.

Anscombe, G. E. M. (2001). *An introduction to Wittgenstein's Tractatus*. St. Augustine's Press. (Original work published 1959)

Balog, K. (2016). Illusionism's discontent. *Journal of Consciousness Studies*, *23*(11–12), 40–51.

Bellet, J., Gay, M., Dwarakanath, A., Jarraya, B., van Kerkoerle, T., Dehaene, S., & Panagiotaropoulos, T. I. (2022). Decoding rapidly presented visual stimuli from prefrontal ensembles without report nor post-perceptual processing. *Neuroscience of Consciousness*, *8*(1). https://doi.org/10.1093/nc/niac005

Blackmore, S., & Troscianko, E. T. (2018). *Consciousness: An introduction* (3rd ed.). Routledge.

Blajenkova, O., Kozhevnikov, M., & Motes, M. A. (2006). Object-spatial imagery: A new self-report imagery questionnaire. *Applied Cognitive Psychology*, *20*(2), 239–263. https://doi.org/10.1002/acp.1182

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

Boly, M., Massimini, M., Tsuchiya, N., Postle, B. R., Koch, C., & Tononi, G. (2017). Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and neuroimaging evidence. *Journal of Neuroscience*, *37*(40), 9603–9613. https://doi.org/10.1523/JNEUROSCI.3218-16.2017

Bornmann, L., Haunschild, R., & Mutz, R. (2021). Growth rates of modern science: A latent piecewise growth curve approach to model publication numbers from established and new literature databases. *Humanities and Social Sciences Communications*, *8*(1), 1–15. https://doi.org/10.1057/s41599-021-00903-w

Bourget, D., & Chalmers, D. J. (2021). Philosophers on philosophy: The 2020 philpapers survey. *Unpublished manuscript*. https://philpapers.org/rec/BOUPOP-3

Brown, R., Lau, H., & LeDoux, J. E. (2019). Understanding the higher-order approach to consciousness. *Trends in Cognitive Sciences*, *23*(9), 754–768. https://doi.org/10.1016/j.tics.2019.06.009

Byrd, N. (2022). Great minds do not think alike: Philosophers' views predicted by reflection, education, personality, and other demographic differences. *Review of Philosophy and Psychology*, 1–38. https://doi.org/10.1007/s13164-022-00628-y

Byrd, N., & Conway, P. (2019). Not all who ponder count costs: Arithmetic reflection predicts utilitarian tendencies, but logical reflection predicts both deontological and utilitarian tendencies. *Cognition*, *192*. https://doi.org/10.1016/j.cognition.2019.06.007

Byrne, A. (2020, November 10). *Inverted qualia*. The Stanford Encyclopedia of Philosophy. Retrieved March 20, 2023, from https://plato.stanford.edu/archives/fall2020/entries/qualia-inverted/

Carruthers, P. (2020). How mindreading might mislead cognitive science. *Journal of Consciousness Studies*, *27*(7–8), 195–219.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, *2*(3), 200–219.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.

Chalmers, D. J. (2018). The meta-problem of consciousness. *Journal of Consciousness Studies*, *25*(9–10), 6–61.

Chalmers, D. J. (2020). Is the hard problem of consciousness universal? *Journal of Consciousness Studies*, *27*(5–6), 227–257.

Cheng, T., Lin, Y., & Tseng, P. (2022). Taking conceptual issues really seriously: One next step for the cognitive science of consciousness. *Cognitive Science*, *46*(11), 1–4. https://doi.org/10.1111/cogs.13213

Clark, C. J., Costello, T., Mitchell, G., & Tetlock, P. E. (2022). Keep your enemies close: Adversarial collaborations will improve behavioral science. *Journal of Applied Research in Memory and Cognition*, *11*(1), 1–18. https://doi.org/10.1037/mac0000004

Clark, C. J., Honeycutt, N., & Jussim, L. (2022). Replicability and the psychology of science. In W. T. O'Donohue, A. Masuda, & S. O. Lilienfeld (Eds.), *Avoiding Questionable Research Practices in Applied Psychology* (pp. 45–71). Springer Cham.

Clark, C. J., & Tetlock, P. E. (2022). Adversarial collaboration: The next science reform. In C. L. Frisby, R. E. Redding, W. T. O'Donohue, & S. O. Lilienfeld (Eds.), *Political Bias in Psychology: Nature, Scope, and Solutions*. Springer.

Cleeremans, A. (2022). Theory as adversarial collaboration. *Nature Human Behaviour*, *6*(4), 485–486. https://doi.org/10.1038/s41562-021-01285-4

Cohen, M. A., & Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends In Cognitive Sciences*, *15*(8), 358–364. https://doi.org/10.1016/j.tics.2011.06.008

Corr, P. J., & Matthews, G. E. (Eds.). (2020). *The Cambridge handbook of personality psychology*. Cambridge University Press.

Dawkins, R. (2015). *The blind watchmaker: Why the evidence of evolution reveals a universe without design*. Penguin Books.

Del Pin, S. H., Skóra, Z., Sandberg, K., Overgaard, M., & Wierzchoń, M. (2021). Comparing theories of consciousness: Why it matters and how to do it. *Neuroscience of Consciousness*, *7*(2), 1–8. https://doi.org/10.1093/nc/niab019

Dennett, D. C. (1987). *The intentional stance*. MIT press.

Dennett, D. C. (2017). *From bacteria to Bach and back: The evolution of minds*. W.W. Norton.

Dennett, D. C. (2018). The fantasy of first-person science. In F. Doria & S. Wuppuluri (Eds.), *The Map and the Territory: Exploring the Foundations of Science, Thought and Reality* (pp. 455–473). Springer Cham.

Descartes, R. (1996). *Descartes: Meditations on first philosophy: With selections from the objections and replies* (J. Cottingham, Ed.; J. Cottingham, Trans.). Cambridge University Press. (Original work published 1641)

Díaz, R. (2021). Do people think consciousness poses a hard problem?: Empirical evidence on the meta-problem of consciousness. *Journal of Consciousness Studies*, *28*(3–4), 55–75.

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. Palgrave Macmillan.

Doerig, A., Schurger, A., & Herzog, M. H. (2021a). Hard criteria for empirical theories of consciousness. *Cognitive Neuroscience*, *12*(2), 41–62. https://doi.org/10.1080/17588928.2020.1772214

Doerig, A., Schurger, A., & Herzog, M. H. (2021b). Response to commentaries on 'hard criteria for empirical theories of consciousness'. *Cognitive Neuroscience*, *12*(2), 99–101. https://doi.org/10.1080/17588928.2020.1853086

Dolan, B. (2007). Soul searching: A brief history of the mind/body debate in the neurosciences. *Neurosurgical Focus*, *23*(1), 1–7. https://doi.org/10.3171/FOC-07/07/E2

Dulany, D. E. (2014). What explains consciousness? Or… What consciousness explains? *Mens Sana Monographs*, *12*(1), 11–34.

Ellenberger, H. F. (1994). *The discovery of the unconscious: The history and evolution of dynamic psychiatry*. Fontana Press.

Esparza Oviedo, S. M. (2020). Similitudes y diferencias en la conceptualización de la conciencia ofrecida por el materialismo eliminativo y el funcionalismo. Un análisis crítico. *Revista Filosofía UIS*, *19*(1), 203–228. https://doi.org/10.18273/revfil.v19n1-2020007

Facco, E., Lucangeli, D., & Tressoldi, P. (2017). On the science of consciousness: Epistemological reflections and clinical implications. *Explore*, *13*(3), 163–180. https://doi.org/10.1016/j.explore.2017.02.007

Fahrenfort, J. J., & van Gaal, S. (2022). Criteria for empirical theories of consciousness should focus on the explanatory power of mechanisms, not on functional equivalence. *Cognitive Neuroscience*, *12*(2), 93–94. https://doi.org/10.1080/17588928.2020.1838470

Faw, B. (2009). Conflicting intuitions may be based on differing abilities: Evidence from mental imaging research. *Journal of Consciousness Studies*, *16*(4), 45–68.

Feist, G. J. (1998). A meta-analysis of personality in scientific and artistic creativity. *Personality and Social Psychology Review*, *2*(4), 290–309.

Feltz, A., & Cokely, E. T. (2009). Do judgments about freedom and responsibility depend on who you are? Personality differences in intuitions about compatibilism and

incompatibilism. *Consciousness and Cognition*, *18*(1), 342–350. https://doi.org/10.1016/j.concog.2008.08.001

Feltz, A., & Cokely, E. T. (2012). The philosophical personality argument. *Philosophical Studies*, *161*(2), 227–246. https://doi.org/10.1007/s11098-011-9731-4

Feltz, A., & Cokely, E. T. (2016). Personality and philosophical bias. In J. M. Sytsma & W. Buckwalter (Eds.), *A Companion to Experimental Philosophy* (pp. 578–589). John Wiley & Sons. https://doi.org/10.1002/9781118661666.ch41

Feltz, A., & Cokely, E. T. (2019). Extraversion and compatibilist intuitions: A ten-year retrospective and meta-analyses. *Philosophical Psychology*, *32*(3), 388–403. https://doi.org/10.1080/09515089.2019.1572692

Fesce, R. (2023). Imagination: The dawn of consciousness: Fighting some misconceptions in the discussion about consciousness. *Physiology & Behavior*, *259*, 1–16. https://doi.org/10.1016/j.physbeh.2022.114035

Fields, C. (2021). What is a theory of consciousness for? *Journal of Consciousness Studies*, *28*(9–10), 104–115.

Fischer, E., & Sytsma, J. M. (2021). Zombie intuitions. *Cognition*, *215*, 1–12. https://doi.org/10.1016/j.cognition.2021.104807

Francken, J. C., Beerendonk, L., Molenaar, D., Fahrenfort, J. J., Kiverstein, J. D., Seth, A. K., & van Gaal, S. (2022). An academic survey on theoretical foundations, common assumptions and the current state of consciousness science. *Neuroscience of Consciousness*, *8*(1), 1–13. https://doi.org/10.1093/nc/niac011

Frankish, K. (2012). Quining diet qualia. *Consciousness and Cognition*, *21*(2), 667–676. https://doi.org/10.1016/j.concog.2011.04.001

Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, *23*(11–12), 11–39.

Frankish, K. (2019). The meta-problem is the problem of consciousness. *Journal of Consciousness Studies*, *26*(9–10), 83–94.

Friedman, D. A., & Søvik, E. (2021). The ant colony as a test for scientific theories of consciousness. *Synthese*, *198*(2), 1457–1480. https://doi.org/10.1007/s11229-019-02130-y

Fürst, M. (2011). What Mary's aboutness is about. *Acta Analytica*, *26*(1), 63–74. https://doi.org/10.1007/s12136-010-0120-y

García-Castro, J. A. (2019). Nuevas teorías sobre la consciencia. *eNeurobiología*, *10*(24). https://www.uv.mx/eneurobiologia/vols/2019/24/Garc%C3%ADa/HTML.html

Gonnerman, C. (2018). Consciousness and experimental philosophy. In R. J. Gennaro (Ed.), *The Routledge Handbook of Consciousness* (1st ed., pp. 463–476). Routledge. https://doi.org/10.4324/9781315676982

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, *26*(2), 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Graziano, M. S., Guterstam, A., Bio, B. J., & Wilterson, A. I. (2020). Toward a standard model of consciousness: Reconciling the attention schema, global workspace, higher-order thought, and illusionist theories. *Cognitive Neuropsychology*, *37*(3–4), 155–172. https://doi.org/10.1080/02643294.2019.1670630

Gregory, D., Hendrickx, M., & Turner, C. (2022). Who knows what Mary knew? An experimental study. *Philosophical Psychology*, *35*(4), 522–545. https://doi.org/10.1080/09515089.2021.2001448

Hanson, J. R., & Walker, S. I. (2021). Formalizing falsification for theories of consciousness across computational hierarchies. *Neuroscience of Consciousness*, *7*(2), 1–11. https://doi.org/10.1093/nc/niab014

Haun, A., & Tsuchiya, N. (2021). Reasonable criteria for functionalists; scarce criteria from phenomenological perspective. *Cognitive Neuroscience*, *12*(2), 95–96. https://doi.org/10.1080/17588928.2020.1838473

Herzog, M. H., Schurger, A., & Doerig, A. (2022). First-person experience cannot rescue causal structure theories from the unfolding argument. *Consciousness and Cognition*, *98*, 1–12. https://doi.org/10.1016/j.concog.2021.103261

Holtzman, G. (2013). Do personality effects mean philosophy is intrinsically subjective? *Journal of Consciousness Studies*, *20*(5–6), 27–42.

Huebner, B. (2010). Commonsense concepts of phenomenal consciousness: Does anyone care about functional zombies? *Phenomenology and the Cognitive Sciences*, *9*(1), 133–155. https://doi.org/10.1007/s11097-009-9126-6

Irvine, E. (2009). Signal detection theory, the exclusion failure paradigm and weak consciousness—Evidence for the access/phenomenal distinction? *Consciousness and Cognition*, *18*(2), 551–560. https://doi.org/10.1016/j.concog.2008.11.002

Irvine, E. (2019). Explaining variation within the meta-problem. *Journal of Consciousness Studies*, *26*(9–10), 115–123.

Jackson, F. (1982). Epiphenomenal qualia. *The Philosophical Quarterly*, *32*(127), 127–136.

Jung, C. G. (2016). *Psychological types*. Routledge. (Original work published 1921)

Kammerer, F. (2016). The hardest aspect of the illusion problem–and how to solve it. *Journal of Consciousness Studies*, *23*(11–12), 124–139.

Kammerer, F. (2021). Certainty and our sense of acquaintance with experiences. *Erkenntnis*, 1–22. https://doi.org/10.1007/s10670-021-00488-5

Kammerer, F. (2022). How can you be so sure? Illusionism and the obviousness of phenomenal consciousness. *Philosophical Studies*, *179*(9), 2845–2867. https://doi.org/10.1007/s11098-022-01804-7

Kant, I. (2005). *Kritikk av den rene fornuft* (S. Mathisen, C. Serck-Hanssen, & Ø. Skar, Trans.). Pax. (Original work published 1781)

Keil, F. C. (1996). The growth of causal understandings of natural kinds. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 234–267). Oxford University Press.

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology*, *57*, 227–254. https://doi.org/10.1146/annurev.psych.57.102904.190100

Kent, L., & Wittmann, M. (2021). Time consciousness: The missing link in theories of consciousness. *Neuroscience of Consciousness*, *7*(2), 1–10. https://doi.org/10.1093/nc/niab011

Kierkegaard, S. (1994). *Avsluttende uvitenskapelig etterskrift til "De philosophiske smuler"* (A. Næss, Ed.). Pax. (Original work published 1846)

Kirkeby-Hinrup, A., & Fazekas, P. (2021). Consciousness and inference to the best explanation: Compiling empirical evidence supporting the access-phenomenal distinction and the overflow hypothesis. *Consciousness and Cognition*, *94*, 1–19. https://doi.org/10.1016/j.concog.2021.103173

Klein, S. B. (2021). Thoughts on the scientific study of phenomenal consciousness. *Psychology of Consciousness: Theory, Research, and Practice*, *8*(1), 74–80. https://doi.org/10.1037/cns0000231

Kleiner, J., & Hoel, E. (2021). Falsification and consciousness. *Neuroscience of Consciousness*, *7*(1), 1–15. https://doi.org/10.1093/nc/niab001

Klempe, S. H. (2008). *Fra opplysning til eksperiment: Om psykologiens oppkomst fra Wolff til Wundt*. Fagbokforlaget.

Klempe, S. H. (2020). The reformation and protestantism's need for psychology. In *Tracing the Emergence of Psychology, 1520–1750: A Sophisticated Intruder to Philosophy* (pp. 59–76). Springer Cham. https://doi.org/10.1007/978-3-030-53701-2_5

Knobe, J., & Prinz, J. (2008). Intuitions about consciousness: Experimental studies. *Phenomenology and the Cognitive Sciences*, *7*(1), 67–83. https://doi.org/10.1007/s11097-007-9066-y

Koch, C. (2004). *The quest for consciousness: A neurobiological approach*. Roberts and Company.

Kozhevnikov, M. (2007). Cognitive styles in the context of modern psychology: Toward an integrated framework of cognitive style. *Psychological Bulletin*, *133*(3), 464–481. https://doi.org/10.1037/0033-2909.133.3.464

Kozhevnikov, M., & Hegarty, M. (2001). Impetus beliefs as default heuristics: Dissociation between explicit and implicit knowledge about motion. *Psychonomic Bulletin & Review*, *8*(3), 439–453. https://doi.org/10.3758/BF03196179

Kuhn, T. S. (2012). *The structure of scientific revolutions* (I. Hacking, Ed.). University of Chicago Press. (Original work published 1962)

Kulstad, M., & Carlin, L. (2020, June 29). *Leibniz's philosophy of mind*. The Stanford Encyclopedia of Philosophy. Retrieved March 20, 2023, from https://plato.stanford.edu/archives/win2020/entries/leibniz-mind/

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Lahav, N., & Neemeh, Z. A. (2022). A relativistic theory of consciousness. *Frontiers in Psychology*, *12*, 1–25. https://doi.org/10.3389/fpsyg.2021.704270

Lamme, V. A. F. (2010). How neuroscience will change our view on consciousness. *Cognitive Neuroscience*, *1*(3), 204–220. https://doi.org/10.1080/17588921003731586

Lamme, V. A. F. (2014). The crack of dawn. In T. K. Metzinger & J. M. Windt (Eds.), *Perceptual Functions and Neural Mechanisms That Mark the Transition from Unconscious Processing to Conscious Vision*. Open MIND. https://doi.org/10.15502/9783958570092

Lamme, V. A. F. (2018). Challenges for theories of consciousness: Seeing or knowing, the missing ingredient and how to deal with panpsychism. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1755), 1–12. https://doi.org/10.1098/rstb.2017.0344

Larsen, R. J., & Buss, D. M. (2017). *Personality psychology: Domains of knowledge about human nature* (6th ed.). McGraw-Hill Education.

Lau, H., & Michel, M. (2019). A socio-historical take on the meta-problem of consciousness. *Journal of Consciousness Studies*, *26*(9–10), 136–147.

Lefcourt, H. M. (1991). Locus of control. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of Personality and Social Psychological Attitudes* (pp. 413–499). Academic Press. https://doi.org/10.1016/B978-0-12-590241-0.50013-7

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, *64*(4), 354–361.

Luczak, A., & Kubo, Y. (2022). Predictive neuronal adaptation as a basis for consciousness. *Frontiers in Systems Neuroscience*, *15*, 1–12. https://doi.org/10.3389/fnsys.2021.767461

Lupyan, G., Uchiyama, R., Thompson, B., & Casasanto, D. (2023). Hidden differences in phenomenal experience. *Cognitive Science*, *47*(1), 1–7. https://doi.org/10.1111/cogs.13239

Mashour, G. A., Roelfsema, P., Changeux, J. P., & Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. *Neuron*, *105*(5), 776–798. https://doi.org/10.1016/j.neuron.2020.01.026

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*(2), 175–215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

Melloni, L., Mudrik, L., Pitts, M., & Koch, C. (2021). Making the hard problem of consciousness easier. *Science*, *372*(6545), 911–912. https://doi.org/10.1126/science.abj3259

Michel, M. (2019). Consciousness science underdetermined: A short history of endless debates. *Ergo, an Open Access Journal of Philosophy*, *6*(28), 771–809. https://doi.org/10.3998/ergo.12405314.0006.028

Mindt, G. (2017). The problem with the 'information' in integrated information theory. *Journal of Consciousness Studies*, *24*(7–8), 130–154.

Miracchi, L. (2019). None of these problems are that 'hard'... or 'easy' making progress on the problems of consciousness. *Journal of Consciousness Studies*, *26*(9–10), 160–172.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, *4*(1), 1–9. https://doi.org/10.1136/bmj.g7647

Naccache, L. (2018). Why and how access consciousness can account for phenomenal consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1755), 1–9. https://doi.org/10.1098/rstb.2017.0357

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, *83*(4), 435–450.

Newman, M., & Gough, D. (2020). Systematic reviews in educational research: Methodology, perspectives and application. In O. Zawacki-Richter, M. Kerres, S. Bedenlier, M. Bond, & K. Buntins (Eds.), *Systematic Reviews in Educational Research* (pp. 3–22). Springer VS. https://doi.org/10.1007/978-3-658-27602-7_1

Nietzsche, F. (2000). Beyond good and evil (W. Kaufmann, Trans.). In *Basic Writings of Nietzsche* (pp. 179–435). Modern Library. (Original work published 1886)

Niikawa, T. (2021). Illusionism and definitions of phenomenal consciousness. *Philosophical Studies*, *178*(1), 1–21. https://doi.org/10.1007/s11098-020-01418-x

Noel, J. P., Faivre, N., Magosso, E., Blanke, O., Alais, D., & Wallace, M. (2019). Multisensory perceptual awareness: Categorical or graded? *Cortex, 120*, 169–180. https://doi.org/10.1016/j.cortex.2019.05.018

Northoff, G., & Lamme, V. A. F. (2020). Neural signs and mechanisms of consciousness: Is there a potential convergence of theories of consciousness in sight? *Neuroscience & Biobehavioral Reviews*, *118*, 568–587. https://doi.org/10.1016/j.neubiorev.2020.07.019

Odegaard, B., Knight, R. T., & Lau, H. (2017). Should a few null findings falsify prefrontal theories of conscious perception? *Journal of Neuroscience*, *37*(40), 9593–9602. https://doi.org/10.1523/JNEUROSCI.3217-16.2017

Oizumi, M., Albantakis, L., & Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: Integrated information theory 3.0. *PLoS Computational Biology*, *10*(5), 1–25. https://doi.org/10.1371/journal.pcbi.1003588

Okasha, S. (2016). *Philosophy of science: A very short Introduction* (2nd ed.). Oxford Paperbacks.

O'Regan, J. K. (2021). Missing: Empirical theories of phenomenal consciousness. *Cognitive Neuroscience*, *12*(2), 82–83. https://doi.org/10.1080/17588928.2020.1838472

Overgaard, M. (2018). Phenomenal consciousness and cognitive access. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1755), 1–6. https://doi.org/10.1098/rstb.2017.0353

Pantani, M., Tagini, A., & Raffone, A. (2018). Phenomenal consciousness, access consciousness and self across waking and dreaming: Bridging phenomenology and neuroscience. *Phenomenology and the Cognitive Sciences*, *17*(1), 175–197. https://doi.org/10.1007/s11097-016-9491-x

Park, J. (2013). The hard problem of consciousness & the progressivism of scientific explanation. *Journal of Consciousness Studies*, *20*(9–10), 90–110.

Peterson, C. (1991). The meaning and measurement of explanatory style. *Psychological Inquiry*, *2*(1), 1–10. https://doi.org/10.1207/s15327965pli0201_1

Pinker, S. (2007). The mystery of consciousness. *Time*, *169*(5), 58–62.

Pinto, Y., & Stein, T. (2021). The hard problem makes the easy problems hard–A reply to Doerig et al. *Cognitive Neuroscience*, *12*(2), 97–98. https://doi.org/10.1080/17588928.2020.1838469

Price, M. C. (1996). Should we expect to feel as if we understand consciousness? *Journal of Consciousness Studies*, *3*(4), 303–312.

Promet, L., & Bachmann, T. (2022). A comparative analysis of empirical theories of consciousness. *Psychology of Consciousness: Theory, Research, and Practice*, 1–34. https://doi.org/10.1037/cns0000341

Raffman, D. (1995). On the persistence of phenomenology. In T. K. Metzinger (Ed.), *Conscious Experience* (pp. 293–308). Ferdinand Schoningh.

Raffone, A., & Pantani, M. (2010). A global workspace model for phenomenal and access consciousness. *Consciousness and Cognition*, *19*(2), 580–596. https://doi.org/10.1016/j.concog.2010.03.013

Rahimian, S. (2022). The myth of when and where: How false assumptions still haunt theories of consciousness. *Consciousness and Cognition*, *97*, 1–7. https://doi.org/10.1016/j.concog.2021.103246

Reisberg, D., Pearson, D. G., & Kosslyn, S. M. (2003). Intuitions and introspections about imagery: The role of imagery experience in shaping an investigator's theoretical views. *Applied Cognitive Psychology*, *17*(2), 147–160. https://doi.org/10.1002/acp.858

Revonsuo, A. (2010). *Consciousness: The science of subjectivity*. Psychology Press.

Revonsuo, A., & Koivisto, M. (2010). Electrophysiological evidence for phenomenal consciousness. *Cognitive Neuroscience*, *1*(3), 225–227. https://doi.org/10.1080/17588928.2010.497580

Riding, R. J. (1997). On the nature of cognitive style. *Educational Psychology*, *17*(1–2), 29–49. https://doi.org/10.1080/0144341970170102

Rosenthal, D. M. (2002). How many kinds of consciousness? *Consciousness and Cognition*, *11*(4), 653–665. https://doi.org/10.1016/S1053-8100(02)00017-X

Rosenthal, D. M. (2021). Assessing criteria for theories. *Cognitive Neuroscience*, *12*(2), 84–85. https://doi.org/10.1080/17588928.2020.1838471

Rosseinsky, N. M. (2019). The troubles with (as-is) consciousness science. *PsyArXiv*. https://doi.org/10.31234/osf.io/nc68y

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied*, *80*(1), 1–28. https://doi.org/10.1037/h0092976

Ruan, Z. (2022). The fundamental challenge of a future theory of consciousness. *Frontiers in Psychology*, *13*, 1–5. https://doi.org/10.3389/fpsyg.2022.1029105

Safron, A. (2020). An integrated world modeling theory (IWMT) of consciousness: Combining integrated information and global neuronal workspace theories with the

free energy principle and active inference framework; toward solving the hard problem and characterizing agentic. *Frontiers in Artificial Intelligence*, *3*, 1–29 https://doi.org/10.3389/frai.2020.00030

Sattin, D., Magnani, F. G., Bartesaghi, L., Caputo, M., Fittipaldo, A. V., Cacciatore, M., Picozzi, M., & Leonardi, M. (2021). Theoretical models of consciousness: A scoping review. *Brain Sciences*, *11*(5), 535. https://doi.org/10.3390/brainsci11050535

Schier, E. (2009). Identifying phenomenal consciousness. *Consciousness and Cognition*, *18*(1), 216–222. https://doi.org/10.1016/j.concog.2008.04.001

Schurger, A., & Graziano, M. (2022). Consciousness explained or described? *Neuroscience of Consciousness*, *8*(1), 1–9. https://doi.org/10.1093/nc/niac001

Seager, W. (2016). *Theories of consciousness: An introduction and assessment* (2nd ed.). Routledge.

Sebastián, M. Á. (2016). Cognitive access and cognitive phenomenology: Conceptual and empirical issues. *Philosophical Explorations*, *19*(2), 188–204. https://doi.org/10.1080/13869795.2016.1176235

Segaller, S. (Director). (2000). *Jung on Film* [Film; VHS]. Public Media Video. (Original work published 1957)

Sękowski, K., & Rorot, W. (2022). Intuition-driven navigation of the hard problem of consciousness. *Review of Philosophy and Psychology*, *13*(1), 239–255. https://doi.org/10.1007/s13164-021-00533-w

Sergent, C., & Rees, G. (2007). Conscious access overflows overt report. *Behavioral and Brain Sciences*, *30*(5–6), 523–524. https://doi.org/10.1017/S0140525X07003044

Seth, A. K., & Bayne, T. (2022). Theories of consciousness. *Nature Reviews Neuroscience*, *23*(7), 439–452. https://doi.org/10.1038/s41583-022-00587-4

Seth, A. K., & Hohwy, J. (2021). Predictive processing as an empirical theory for consciousness science. *Cognitive Neuroscience*, *12*(2), 89–90. https://doi.org/10.1080/17588928.2020.1838467

Shenhav, A., Rand, D. G., & Greene, J. D. (2012). Divine intuition: Cognitive style influences belief in God. *Journal of Experimental Psychology: General*, *141*(3), 423–428. https://doi.org/10.1037/a0025391

Signorelli, C. M., Cea, I., & Prentner, R. (2022). We need to explain subjective experience, but its explanation may not be mechanistic. *PsyArXiv*. https://doi.org/10.31234/osf.io/e6kdg

Signorelli, C. M., Szczotka, J., & Prentner, R. (2021). Explanatory profiles of models of consciousness–Towards a systematic classification. *Neuroscience of Consciousness*, *7*(2), 1–13. https://doi.org/10.1093/nc/niab021

Simonton, D. K. (2009). Varieties of (scientific) creativity: A hierarchical model of domain-specific disposition, development, and achievement. *Perspectives on Psychological Science*, *4*(5), 441–452.

Stoljar, D. (2016). The semantics of "what it's like" and the nature of consciousness. *Mind*, *125*(500), 1161–1198. https://doi.org/10.1093/mind/fzv179

Sytsma, J. M., & Machery, E. (2009). How to study folk intuitions about phenomenal consciousness. *Philosophical Psychology*, *22*(1), 21–35. https://doi.org/10.1080/09515080802703653

Sytsma, J. M., & Ozdemir, E. (2019). No problem: Evidence that the concept of phenomenal consciousness is not widespread. *Journal of Consciousness Studies*, *26*(9–10), 241–256.

Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience*, *17*(7), 450–461. https://doi.org/10.1038/nrn.2016.44

Tye, M. (1999). Phenomenal consciousness: The explanatory gap as a cognitive illusion. *Mind*, *108*(432), 705–725. https://doi.org/10.1093/mind/108.432.705

Usher, M., Bronfman, Z. Z., Talmor, S., Jacobson, H., & Eitam, B. (2018). Consciousness without report: Insights from summary statistics and inattention 'blindness'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1755), 1–11. https://doi.org/10.1098/rstb.2017.0354

Van Gulick, R. (1994). Dennett, drafts, and phenomenal realism. *Philosophical Topics*, *22*(1/2), 443–455. https://doi.org/10.5840/philtopics1994221/21

Varela, F. J. (1996). Neurophenomenology: A methodological remedy for the hard problem. *Journal of Consciousness Studies*, *3*(4), 330–349.

Velmans, M. (1996). *The Science of Consciousness: Psychological, Neuropsychological and Clinical Reviews*. Routledge.

Walter, N., & Hinterberger, T. (2022). Self-organized criticality as a framework for consciousness: A review study. *Frontiers in Psychology*, *13*, 1–12. https://doi.org/10.3389/fpsyg.2022.911620

White, B. (2021). The hard problem isn't getting any easier: Thoughts on Chalmers' "meta-problem". *Philosophia*, *49*(1), 495–506. https://doi.org/10.1007/s11406-020-00210-9

Wiese, W. (2020). The science of consciousness does not need another theory, it needs a minimal unifying model. *Neuroscience of Consciousness*, *6*(1), 1–7. https://doi.org/10.1093/nc/niaa013

Winters, J. J. (2020). The temporally-integrated causality landscape: A theoretical framework for consciousness and meaning. *Consciousness and Cognition*, *83*, 1–14. https://doi.org/10.1016/j.concog.2020.102976

Winters, J. J. (2021). The temporally-integrated causality landscape: Reconciling neuroscientific theories with the phenomenology of consciousness. *Frontiers in Human Neuroscience*, *15*, 1–14. https://doi.org/10.3389/fnhum.2021.768459

Wittgenstein, L. (2009). *Philosophical investigations* (P. S. Hacker & J. Schulte, Eds.). Wiley. (Original work published 1953)

Wu, W. (2018, October 9). *The neuroscience of consciousness*. Stanford Encyclopedia of Philosophy. Retrieved February 20, 2023, from https://plato.stanford.edu/archives/win2018/entries/consciousness-neuroscience/

Wyrwa, M. (2022). Does the folk concept of phenomenal consciousness exist? *Diametros*, *19*(71), 46–66. https://doi.org/10.33392/diam.1751

Yaden, D. B., & Anderson, D. E. (2021). The psychology of philosophy: Associating philosophical views with psychological traits in professional philosophers. *Philosophical Psychology*, *34*(5), 721–755. https://doi.org/10.1080/09515089.2021.1915972

Yaron, I., Melloni, L., Pitts, M., & Mudrik, L. (2021). The consciousness theories studies (ConTraSt) database: Analyzing and comparing empirical studies of consciousness theories. *bioRxiv*. https://doi.org/10.1101/2021.06.10.447863

Yeung, A. W. K., Cushing, C. A., & Lee, A. L. F. (2022). A bibliometric evaluation of the impact of theories of consciousness in academia and on social media. *Consciousness and Cognition*, *100*, 1–14. https://doi.org/10.1016/j.concog.2022.103296

Yurchenko, S. B. (2022). From the origins to the stream of consciousness and its neural correlates. *Frontiers in Integrative Neuroscience*, *16*, 1–26. https://doi.org/10.3389/fnint.2022.928978

Zeman, A., Milton, F., Della Sala, S., Dewar, M., Frayling, T., Gaddum, J., Hattersley, A., Heuerman-Williamson, B., Jones, K., MacKisack, M., & Winlove, C. (2020). Phantasia–The psychological significance of lifelong visual imagery vividness extremes. *Cortex*, *130*, 426–440. https://doi.org/10.1016/j.cortex.2020.04.003

Appendix A

Search strings:

The database used and applied criteria outside of the search string have been put in brackets.

RQ1: [Google Scholar; years 2019-2022; first 20 results] "Theories of consciousness" AND "Models of consciousness"

RQ2: [WoS; years 2019-2022] (((TS=(theor* NEAR/2 conscious*)) OR TS=(model* NEAR/2 conscious*)) AND ALL=(neuro*)) AND AB=(Consensus OR Accept* NEAR/5 theor* OR Agree* OR Converg* NEAR/5 theor* OR Discrepanc* OR Divers* NEAR/5 theor* OR Heterogen*)

RQ3:

1. [WoS] TS=((neuro* NEAR/5 theor* NEAR/2 conscious*) OR (neuro* NEAR/5 model* NEAR/2 conscious*))
2. [WoS] TS=((neuro* NEAR/5 theor* NEAR/2 conscious*) OR (neuro* NEAR/5 model* NEAR/2 conscious*)) AND ALL=((explanatory AND gap) OR (hard AND problem))
3. [WoS] TS=((neuro* NEAR/5 theor* NEAR/2 conscious*) OR (neuro* NEAR/5 model* NEAR/2 conscious*)) AND ALL=(phenomenolog* OR phenomenal*)
4. [WoS] ALL=(A)
5. [WoS; WoS Neurosciences] ALL=(A)
6. [WoS] ALL=(Consciousness)

RQ4: [WoS] ALL=(phenomenal) AND ALL=(access) AND ALL=(conscious*) AND ALL=(theor*) AND ALL=(neurosci*)

RQ5:

1. [Google Search; first 5 pages] Chalmers + Hard Problem + Encyclopedia
2. [Google Search; first 5 pages] Levine + Explanatory Gap + Encyclopedia [*no quotations found*]
3. [Google Search; first 5 pages] Descartes + Mind-Body + Encyclopedia
4. [Google Search; first 5 pages] Leibniz + Mill + Encyclopedia
5. [Google Search; first 5 pages] Jackson + Mary + Encyclopedia
6. [Google Search; first 5 pages] Nagel + Bat + Encyclopedia [*no quotations found*]
7. [Google Search; first 5 pages] Chalmers + Zombie + Encyclopedia [*no quotations found*]
8. [Google Search; first 5 pages] Locke + Inverted + Encyclopedia

Glossary:

"What it is like" / "something it is like"
: Nagel's concept for defining consciousness. The intrinsic state in which (at least) a living creature finds itself when it experiences something.

Hard Problem of consciousness
: Chalmers' concept for demonstrating the difficulties in explaining consciousness. The fact that explaining consciousness in physical terms is seemingly impossible, and thus much harder than other types of explanations.

Explanatory gap
: Levine's concept for demonstrating the difficulties in explaining consciousness. The fact that it seems impossible to explain consciousness in physical terms, leaving a gap between consciousness and the physical terms we use to explain it.

Mind-body problem
: Descartes' problem for demonstrating the difficulties in relating the material world to the immaterial soul. There is no tangible connection which relates the body (e.g., the brain) to the mind (e.g., thought).

Physical states
: States which are describable in physical terms, that is, objective or third-person states. The current organization of some physical matter. For example the current atomic state of a cup, or the current neural state of a brain. Often interchangeable with functional states.

Phenomenal states
: States which are describable in mental terms, that is, subjective or first-person states. The current organization of some subjective experience. For example, the current state of a thought, or the current state of a visual experience.

Phenomenal concepts
: Concepts which aim to describe phenomenal states. For example, phenomenality, acquaintance, transparency, qualia.

Phenomenal consciousness
: Alternatively, "P-consciousness". A concept which aims to describe consciousness in terms of phenomenal concepts. Typically describes intrinsically having consciousness. For example, consciousness as being "what it is like" to have an experience.

Access consciousness
: Alternatively, "A-consciousness". A concept which aims to describe consciousness without referring to phenomenal concepts. Typically describes the contrast between conscious and unconscious states. For example, consciousness as "making information available".

Phenomenal realism
: The broad description of a camp or position in philosophy and consciousness research. In other literature, other versions of "realism" are used. Typically regards the hard problem as a real problem. Focuses on consciousness in terms of its first-person aspects.

Illusionism
: The broad description of a camp or position in philosophy and consciousness research. Typically disregards the hard problem as a real problem. Focuses on consciousness in terms of its third-person aspects.

Problem intuitions

Alternatively, "hard problem intuitions". Our intuitions or introspective recognition that there is a hard problem of consciousness. Typically induced via philosophical thought experiments such as "Mary the Color Scientist".

Meta-problem

Chalmers' concept for demonstrating a psychological component of the hard problem. The problem of why we think that there is a hard problem of consciousness.

Explanandum

The phenomenon which is to be explained in an explanation. The explanatory target.

A priori

Beforehand. A concept which designates something preceding experience or a given fact. An example of a priori knowledge is mathematical deduction: Knowledge which is true regardless of individual experiences.

A posteriori

Afterward. A concept which designates something following experience or a given fact. An example of a posteriori knowledge is empirical induction: Knowledge which follows from individual experiences.

Epistemology

The philosophical inquiry into what knowledge is. Includes investigations of concepts such as truth, justification, belief, and skepticism.

Ontology

The philosophical inquiry into what being is. Includes investigations of concepts such as reality, existence, becoming, substance, and relations.

Theory-ladenness

Concept popularized by Kuhn. Empirical observations are affected by the theoretical presuppositions of the observer. Data are "laden" with theory.

Appendix B

Quotations:

RQ2: Clear lack of consensus (gray), vague lack of consensus or emerging consensus (clear).

Author(s)	Article title	Quotation
Doerig et al. (2021b)	Response to commentaries on 'hard criteria for empirical theories of consciousness'	"Overall, there seems to be consensus that a theory of consciousness (ToC) needs to have an unconscious alternative, but other criteria sparked controversy. The hottest debate is to what extent consciousness needs to work with purely 1(st) person data, containing information not available in 3(rd) person reports."
Esparza Oviedo (2020)	Similitudes y diferencias en la conceptualización de la conciencia ofrecida por el materialismo eliminativo y el funcionalismo. Un análisis crítico	"… it will be argued that in the field of cognitive science there is no consensus, nor a theoretical and conceptual basis on what consciousness is."
Michel (2019)	Consciousness science underdetermined: A short history of endless debates	"Consciousness scientists have not reached consensus on two of the most central questions in their field: first, on whether consciousness overflows reportability; second, on the physical basis of consciousness."
Winters (2021)	The temporally-integrated causality landscape: Reconciling neuroscientific theories with the phenomenology of consciousness	"While the literature concerned with these theories tends to focus on different lines of evidence, there are fundamental areas of agreement. This means that, in time, it may be possible for many of them to converge upon the truth."
Walter and Hinterberger (2022)	Self-organized criticality as a framework for consciousness: A review study	"No current model of consciousness is univocally accepted on either theoretical or empirical grounds, and the need for a solid unifying framework is evident."
Winters (2020)	The temporally-integrated causality landscape: A theoretical framework for consciousness and meaning	"Theoretical approaches to understanding consciousness have begun to converge upon areas of general agreement, yet substantive differences remain."
García-Castro (2019)	Nuevas teorías sobre la consciencia	"There is no agreement at the present moment concerning the concept of consciousness. The

		great diversity of theories, some of them antagonistic, should be reflecting an immature, emerging science about consciousness."
Friedman and Søvik (2021)	The ant colony as a test for scientific theories of consciousness	"Absent an agreed-upon definition of consciousness or even a convenient system to test theories of consciousness, a confusing heterogeneity of theories proliferate."
Kent and Wittmann (2021)	Time consciousness: The missing link in theories of consciousness	"There are plenty of issues to be solved in order for researchers to agree on a neural model of consciousness."
Herzog et al. (2022)	First-person experience cannot rescue causal structure theories from the unfolding argument	"We recently put forward an argument, the Unfolding Argument (UA), that integrated information theory (IIT) and other causal structure theories are either already falsified or unfalsifiable, which provoked significant criticism. It seems that we and the critics agree that the main question in this debate is whether first-person experience, independent of third-person data, is a sufficient foundation for theories of consciousness."
Haun and Tsuchiya (2021)	Reasonable criteria for functionalists; scarce criteria from phenomenological perspective	"… if the field can agree to a family of paradigm cases for consciousness, this would be an important endeavor for the field."
Del Pin et al. (2021)	Comparing theories of consciousness: Why it matters and how to do it	"Nonetheless, when we surveyed publications on consciousness research, we found that most focused on a single theory. When 'comparisons' happened, they were often verbal and non-systematic. This fact in itself could be a contributing reason for the lack of convergence between theories in consciousness research."
Signorelli et al. (2021)	Explanatory profiles of models of consciousness–Towards a systematic classification	"In particular, we argue that different models explicitly or implicitly subscribe to different notions of what constitutes a satisfactory explanation, use different tools in their explanatory endeavours and even aim to explain very different phenomena. We thus present a framework to compare existing models in the field with respect to what we call their 'explanatory profiles'. We focus on the following minimal dimensions: mode of explanation, mechanisms of explanation and target of explanation. We also discuss the empirical consequences of the discussed

		discrepancies among models."
Northoff and Lamme (2020)	Neural signs and mechanisms of consciousness: Is there a potential convergence of theories of consciousness in sight?	"Various theories for the neural basis of consciousness have been proposed, suggesting a diversity of neural signs and mechanisms. We ask to what extent this diversity is real, or whether many theories share the same basic ideas with a potential for convergence towards a more unified theory of the neural basis of consciousness."
Sattin et al. (2021)	Theoretical models of consciousness: A scoping review	"We found heterogeneous perspectives in the theories analyzed. Those with the highest grade of variability are as follows: subjectivity, NCC, and the consciousness/cognitive function."
Luczak and Kubo (2022)	Predictive neuronal adaptation as a basis for consciousness	"Interestingly, our predictive adaptation hypothesis is consistent with multiple ideas presented previously in diverse theories of consciousness, such as global workspace theory, integrated information, attention schema theory, and predictive processing framework. In summary, we present a theoretical, computational, and experimental support for the hypothesis that neuronal adaptation is a possible biological mechanism of conscious processing, and we discuss how this could provide a step toward a unified theory of consciousness."
Doerig et al. (2021a)	Hard criteria for empirical theories of consciousness	"Consciousness is now a well-established field of empirical research. A large body of experimental results has been accumulated and is steadily growing. In parallel, many Theories of Consciousness (ToCs) have been proposed. These theories are diverse in nature, ranging from computational to neurophysiological and quantum theoretical approaches."

RQ4: Well-defined division (gray), questioning whether there is a real distinction (clear).

Author(s)	**Article title**	**Quotation**
Overgaard (2018)	Phenomenal consciousness and cognitive access	"In consciousness research, it is common to distinguish between phenomenal consciousness and access consciousness. Recently, a number of scientists have attempted to show that phenomenal content can be empirically separated from cognitive access and,

		accordingly, that the mental content that is accessed is not (always) identical to the content that is experienced."
Raffone and Pantani (2010)	A global workspace model for phenomenal and access consciousness	"Both the global workspace theory and Block's distinction between phenomenal and access consciousness, are central in the current debates about consciousness and the neural correlates of consciousness."
Schier (2009)	Identifying phenomenal consciousness	"This paper examines the possibility of finding evidence that phenomenal consciousness is independent of access."
O'Regan (2021)	Missing: Empirical theories of phenomenal consciousness	"Doerig et al. evaluate how current empirical theories approach access consciousness, but they neglect how they approach phenomenal consciousness – probably because most theories don't deal with phenomenal consciousness at all."
Naccache (2018)	Why and how access consciousness can account for phenomenal consciousness	"According to a popular distinction proposed by the philosopher Ned Block in 1995, our conscious experience would overflow the very limited set of what we can consciously report to ourselves and to others. He proposed to coin this limited consciousness 'Access Consciousness' (A-Cs) and to define 'Phenomenal Consciousness' as a much richer subjective experience that is not accessed but that would still delineate the extent of consciousness. In this article, I review and develop five major problems raised by this theory, and show how a strict A-Cs theory can account for our conscious experience. I illustrate such an A-Cs account …"
Irvine (2009)	Signal detection theory, the exclusion failure paradigm and weak consciousness—Evidence for the access/phenomenal distinction?	"[Researchers] claim that a signal detection theory (SDT) analysis of qualitative difference paradigms, in particular the exclusion failure paradigm, reveals cases of phenomenal consciousness without access consciousness. This claim is unwarranted on several grounds."
Pantani et al. (2018)	Phenomenal consciousness, access consciousness and self across waking and dreaming: Bridging phenomenology and neuroscience	"The distinction between phenomenal and access consciousness is central to debates about consciousness and its neural correlates."

Kirkeby-Hinrup and Fazekas (2021)	Consciousness and inference to the best explanation: Compiling empirical evidence supporting the access-phenomenal distinction and the overflow hypothesis	"… we deliver a complete collection (the compilation step) of empirical support for the distinction between A-Consciousness and P-Consciousness and the overflow hypothesis."
Revonsuo and Koivisto (2010)	Electrophysiological evidence for phenomenal consciousness	"Overall, the ERP evidence supports the view that phenomenal consciousness of a visual stimulus emerges earlier than access consciousness, and that attention and awareness are served by distinct neural processes."
Sergent and Rees (2007)	Conscious access overflows overt report	"Block proposes that phenomenal experience overflows conscious access. In contrast, we propose that conscious access overflows overt report. We argue that a theory of phenomenal experience cannot discard subjective report and that Block's examples of phenomenal "overflow" relate to two different types of perception. We propose that conscious access is more than simply readout of a preexisting phenomenal experience."
Lamme (2018)	Challenges for theories of consciousness: Seeing or knowing, the missing ingredient and how to deal with panpsychism	"Controversy about whether the conscious experience is better explained by theories that focus on phenomenal (P-consciousness) or cognitive aspects (A-consciousness) remains, and the debate seems to reach a stalemate. Can we ever resolve this?"
Sebastián (2016)	Cognitive access and cognitive phenomenology: Conceptual and empirical issues	"The well-known distinction between access consciousness and phenomenal consciousness has moved away from the conceptual domain into the empirical one, and the debate now is focused on whether the neural mechanisms of cognitive access are constitutive of the neural correlate of phenomenal consciousness. … If the mechanisms responsible for cognitive access can be disentangled from the mechanisms that give rise to phenomenology in the case of perception and emotion, then the same disentanglement is to be expected in the case of thoughts. This, in turn, presents, as I argue, a challenge to the cognitive phenomenology thesis: either there are thoughts with cognitive phenomenology we lack cognitive access to or there are good

		reasons to doubt that there is such a thing as cognitive phenomenology.”
Rosenthal (2002)	How many kinds of consciousness?	“Ned Block's influential distinction between phenomenal and access consciousness has become a staple of current discussions of consciousness. It is not often noted, however, that his distinction tacitly embodies unargued theoretical assumptions that favor some theoretical treatments at the expense of others.”
Usher et al. (2018)	Consciousness without report: Insights from summary statistics and inattention 'blindness'	“We contrast two theoretical positions on the relation between phenomenal and access consciousness. First, we discuss previous data supporting a mild Overflow position, according to which transient visual awareness can overflow report. These data are open to two interpretations: (i) observers transiently experience specific visual elements outside attentional focus without encoding them into working memory; (ii) no specific visual elements but only statistical summaries are experienced in such conditions.”
Block (2005)	Two neural correlates of consciousness	“Neuroscientists continue to search for 'the' neural correlate of consciousness (NCC). In this article, I argue that a framework in which there are at least two distinct NCCs is increasingly making more sense of empirical results than one in which there is a single NCC. I outline the distinction between phenomenal NCC and access NCC, and show how they can be distinguished by experimental approaches, in particular signal-detection theory approaches. Recent findings in cognitive neuroscience provide an empirical case for two different NCCs.”
Noel et al. (2019)	Multisensory perceptual awareness: Categorical or graded?	“Neural evidence suggests that mechanisms associated with conscious access (i.e., the ability to report on a conscious state) are “all-or-none”. Upon crossing some threshold, neural signals are globally broadcast throughout the brain and allow conscious reports. However, whether subjective experience (phenomenal consciousness) is categorical (i.e., transitioning abruptly from unconscious to conscious states) or graded (i.e., characterized by multiple intermediate states) remains an open question.”

Bellet et al. (2022)	Decoding rapidly presented visual stimuli from prefrontal ensembles without report nor post-perceptual processing	"We discuss whether the observed activation reflects conscious access, phenomenal consciousness, or merely a preconscious bottom-up wave."

RQ5:
Classical texts.
Concepts rated as epistemologically laden are highlighted in gray.

Author(s)	**Concept**	**Encyclopedia / article title**	**Quotation**
Chalmers (in online encyclopedia)	The Hard Problem of consciousness	*Internet Encyclopedia of Philosophy:* Hard Problem of Consciousness	"What makes the hard problem hard and almost unique is that it goes beyond problems about the performance of functions. To see this, note that even when we have explained the performance of all the cognitive and behavioral functions in the vicinity of experience—perceptual discrimination, categorization, internal access, verbal report—there may still remain a further unanswered question: *Why is the performance of these functions accompanied by experience?*" (emphasis in original)
Levine (1983, p. 357)	The Explanatory Gap	Materialism and Qualia: The Explanatory Gap	"… there is more to our concept of pain than its causal role, there is its qualitative character, how it feels; and what is left unexplained by the discovery of C-fiber firing is *why pain should feel the way it does!* For there seems to be nothing about C-fiber firing which makes it naturally "fit" the phenomenal properties of pain, any more than it would fit some other set of phenomenal properties. Unlike its functional role, the identification of the qualitative side of pain with C-fiber firing (or some property of C-fiber firing) leaves the connection between it and what we identify it with completely mysterious. One might say, it makes the way pain feels into merely a brute fact." (emphasis in original)

Descartes (in online encyclopedia)	Mind-Body Distinction	*Internet Encyclopedia of Philosophy:* The Mind-Body Distinction	"[T]here is a great difference between the mind and the body, inasmuch as the body is by its very nature always divisible, while the mind is utterly indivisible. For when I consider the mind, or myself in so far as I am merely a thinking thing, I am unable to distinguish any parts within myself; I understand myself to be something quite single and complete….By contrast, there is no corporeal or extended thing that I can think of which in my thought I cannot easily divide into parts; and this very fact makes me understand that it is divisible. This one argument would be enough to show me that the mind is completely different from the body…."
Leibniz (in online encyclopedia)	The windmill metaphor	*Stanford Encyclopedia of Philosophy:* Leibniz's Philosophy of Mind	"One is obliged to admit that *perception* and what depends upon it is *inexplicable on mechanical principles*, that is, by figures and motions. In imagining that there is a machine whose construction would enable it to think, to sense, and to have perception, one could conceive it enlarged while retaining the same proportions, so that one could enter into it, just like into a windmill. Supposing this, one should, when visiting within it, find only parts pushing one another, and never anything by which to explain a perception. Thus it is in the simple substance, and not in the composite or in the machine, that one must look for perception." (emphasis in original)
Jackson (in online encyclopedia)	Mary the Color Scientist	*Stanford Encyclopedia of Philosophy:* Qualia: The Knowledge Argument	"Mary is a brilliant scientist who is, for whatever reason, forced to investigate the world from a black and white room via a black and white television monitor. She specializes in the neurophysiology of vision and acquires, let us suppose, all the physical information there is to obtain about what goes on when we see ripe tomatoes, or the sky, and use terms

			like ‘red’, ‘blue’, and so on. She discovers, for example, just which wavelength combinations from the sky stimulate the retina, and exactly how this produces *via* the central nervous system the contraction of the vocal chords and expulsion of air from the lungs that results in the uttering of the sentence ‘The sky is blue’… What will happen when Mary is released from her black and white room or is given a color television monitor? Will she *learn* anything or not? It seems just obvious that she will learn something about the world and our visual experience of it. But then is it inescapable that her previous knowledge was incomplete. But she had *all* the physical information. *Ergo* there is more to have than that, and Physicalism is false.” (emphasis in original)
Nagel (1974, p. 437)	Phenomenal features	What Is It Like to Be a Bat?	“While an account of the physical basis of mind must explain many things, this appears to be the most difficult. It is impossible to exclude the phenomenological features of experience from a reduction in the same way that one excludes the phenomenal features of an ordinary substance from a physical or chemical reduction of it—namely, by explaining them as effects on the minds of human observers. If physicalism is to be defended, the phenomenological features must themselves be given a physical account. But when we examine their subjective character it seems that such a result is impossible. The reason is that every subjective phenomenon is essentially connected with a single point of view, and it seems inevitable that an objective, physical theory will abandon that point of view.”
Chalmers (1996, p. 84)	Philosophical Zombies	The Conscious Mind: In Search of a Fundamental	“The most obvious way (although not the only way) to investigate the logical supervenience of

		Theory	consciousness is to consider the logical possibility of a zombie: someone or something physically identical to me (or to any other conscious being), but lacking conscious experiences altogether. At the global level, we can consider the logical possibility of a zombie world: a world physically identical to ours, but in which there are no conscious experiences at all. In such a world, everybody is a zombie."
Locke (in online encyclopedia)	Inverted Qualia	*Stanford Encyclopedia of Philosophy:* Inverted Qualia	"Neither would it carry any Imputation of *Falshood* to our simple *Ideas*, if by the different Structure of our Organs, it were so ordered, That *the same Object should produce in several Men's Minds different* Ideas at the same time; *v.g.* if the *Idea*, that a *Violet* produced in one Man's Mind by his Eyes, were the same that a *Marigold* produces in another Man's, and *vice versâ*. For since this could never be known: because one Man's Mind could not pass into another Man's Body, to perceive, what Appearances were produced by those Organs; neither the *Ideas* hereby, nor the Names, would be at all confounded, or any *Falshood* be in either. For all Things, that had the Texture of a *Violet*, producing constantly the *Idea*, which he called *Blue*, and those which had the Texture of a *Marigold*, producing constantly the *Idea*, which he as constantly called *Yellow*, whatever those Appearances were in his Mind; he would be able as regularly to distinguish Things for his Use by those Appearances, and understand, and signify those distinctions, marked by the Names *Blue* and *Yellow*, as if the Appearances, or *Ideas* in his Mind, received from those two Flowers, were exactly the same, with the *Ideas* in other Men's Minds." (emphasis in original)

Theoretical texts.
Concepts rated as epistemologically laden are highlighted in gray.

Author(s)	**Theory**	**Article title**	**Quotation**
Oizumi et al. (2014)	Integrated Information Theory (IIT)	From the phenomenology to the mechanisms of consciousness: Integrated information theory 3.0	"It must be emphasized that taking the phenomenology of consciousness as primary, and asking how it can be implemented by physical mechanisms, is the opposite of the approach usually taken in neuroscience: start from neural mechanisms in the brain, and ask under what conditions they give rise to consciousness, as assessed by behavioral reports. While identifying the "neural correlates of consciousness" is undoubtedly important, it is hard to see how it could ever lead to a satisfactory explanation of what consciousness is and how it comes about."
Oizumi et al. (2014)	Integrated Information Theory (IIT)	From the phenomenology to the mechanisms of consciousness: Integrated information theory 3.0	"One can now ask, for any set of physical mechanisms, whether it is associated with phenomenology (is there "something it is like to be it," from its own intrinsic perspective), how much of it (the quantity or level of consciousness), and of which kind (the quality or content of the experience). As also indicated by the figure, these phenomenological properties should be considered as intrinsic properties of physical mechanisms arranged in a certain way, meaning that a complex of physical mechanisms in a certain state is necessarily associated with its quale."
Tononi et al. (2016)	Integrated Information Theory (IIT)	Integrated information theory: From consciousness to its physical substrate	"The reason why some neural mechanisms, but not others, should be associated with consciousness has been called 'the hard problem' because it seems to defy the possibility of a scientific explanation. In this Opinion article, we provide an overview of the integrated information theory (IIT) of consciousness, which has been developed over the past few years. IIT addresses the hard problem in a new way. It does not start from the brain and ask how it could give rise to experience;

			instead, it starts from the essential phenomenal properties of experience, or axioms, and infers postulates about the characteristics that are required of its physical substrate."
Brown et al. (2019)	Higher-Order Theory (HOT)	Understanding the higher-order approach to consciousness	"Consciousness … as used here, refers to subjective experience, or what is sometimes called phenomenal consciousness, as opposed to the condition of merely being awake and alert and behaviorally responsive to external stimuli. To be phenomenally conscious is for there to be something that it is like to be the entity in question, that is, something that it is like for the entity itself. … Subjective experience is the stuff of novels, poems, and songs, of our emotions and memories, the essence of being a human. It is hard to imagine what it would be like to not be sentient in the way we are. Unsurprisingly, then, the science of consciousness is currently a vibrant and thriving area of research. However, there is no generally accepted theory of the phenomena being studied, and the phenomena themselves often do not include many of the kinds of complex experiences that we normally have in the course of day-to-day life, such as of our emotions and memories."
Lamme (2010)	Recurrent Processing Theory (RPT)	How neuroscience will change our view on consciousness	"Functions, whether cognitive or not, are of course also seen as irrelevant to consciousness in the original formulation of the so called *hard problem* of consciousness. I am not implying that that line of reasoning should be followed fully, as that way of posing the problem makes phenomenality—or qualia—almost impossible to study. For example, it renders invalid the very intuitions on which the conscious–unconscious divide is based … But I do agree that many functions—cognitive functions in particular—do very little towards explaining qualia."
Lamme (2010)	Recurrent Processing	How neuroscience will change our view	"… by linking consciousness so much with cognition, there is some "throwing

	Theory (RPT)	on consciousness	away of the baby with the bathwater," because cognition and access do very little to explain the key feature of consciousness that we consider here, which is phenomenality. Why would combining visual input with working memory make it "visible" to the mind's eye—in other words, produce qualia?"
Lamme (2014)	Recurrent Processing Theory (RPT)	The crack of dawn	"There is one big difference between the camera and the human mind, though. The camera does not see. I do. … It is this aspect of visual processing that is in need of an explanation. Not the fact that I recognize the person in front of me, can read his emotions, talk to him, or pick up the cup of coffee he gives me. I can vaguely understand how my brain enables me to do that. What I do not understand is how it is that I see all those things. … Let's find the visual functions and neural processes that take us as close as possible to the hard problem, as close as possible towards explaining why we humans see, while photo cameras do not."