Axiomathes
https://doi.org/10.1007/s10516-021-09570-7

ORIGINAL PAPER

A Skeptical View on the Physics-Consciousness Explanatory Gap

Mario Martinez-Saito[1]

Received: 7 February 2021 / Accepted: 9 June 2021

Abstract
The epistemological chasm between how we (implicitly and subjectively) perceive or imagine the actual world and how we (explicitly and "objectively") think of its underlying entities has motivated perhaps the most disconcerting impasse in human thought: the explanatory gap between the phenomenal and physical properties of the world. Here, I advocate a combination of philosophical skepticism and simplicity as an informed approach to arbitrate among theories of consciousness. I argue that the explanatory gap is rightly a gap in our understanding, but one that is not surprising; and we being observers biased by our first-person perspective and our existence may both hinder and (the realization we have them) assist our reasoning. Further, I unfold the concept of observer into two distinct notions based on its functional and phenomenal aspects, and exploit this device to elucidate the subject-observer relationship. Then, I proceed to analyze the philosophical zombie dilemma. Lastly, I contend that from a skeptical viewpoint, panpsychism (or neutral monism) is the most parsimonious doctrine accounting for the explanatory gap, and suggest that it would be possible to make headway in the hard problem of consciousness by uncovering non-trivial causal relationships between qualia states and functional states, if routine and controlled manipulation of neural circuits were easily available.

Keywords Occam's razor · Skepticism · Solipsism · Panpsychism · Qualia

1 Introduction: What is Going on Out There?

Is there anything at all out there? By "out there" I am referring to the "outside" of mind, the (hypothetical) external world that we often assume to be "seeing directly"—which we are not. Although many have attempted to answer this question, there is no consensus about what is the right answer, or even about whether there is a right answer. This article presents my thesis on this topic.

✉ Mario Martinez-Saito
mmartinezsaito@gmail.com

[1] Institute of Cognitive Neuroscience, HSE University, Moscow, Russia

Published online: 16 June 2021

Perhaps the most fundamental question in metaphysics is the ontological status of the world out there, usually called the actual world. Most views posit reality to lie somewhere, on the axis aligned to ontology (what the world is made of), between the fundamentally mental (idealism) and the fundamentally material (materialism, including physicalism; Fig. 1, horizontal axis). Another distinction can be established on the basis of the axis (Fig. 1, vertical) defined by dependence on a subjective perspective (e.g. logical positivism, subjective idealism), and independence from the subject, i.e. realism (e.g., philosophical realism, Platonic realism). Realism assumes that there are objects out there in a manner independent of any observer, and thereby objects are granted a fundamental and irreducible ontological role. By contrast, subjectivism assumes that objects exist only or primarily inside the mind of an observer, and thus it lays in the observer some ontological responsibility in understanding reality by judiciously using introspection and logical reasoning. We can make yet another distinction standing on epistemological grounds, inspired by what is perhaps Kant's (1787/1998) greatest contribution to metaphysics: transcendental idealism, and the insight that what things are (noumena, in Kantian jargon, the objects or causes of the world that we assume to be "out there") and how they appear to us (phenomena) are not only two distinct (and plausible) notions, but that this crack in the chain of perception calls for ascribing to perception a fundamentally subject-based quality—with the mental medium enacting a "reality filter" that delivers the phenomena in a baffling and non-intuitive manner. Hence the dependence of the world on the mind can be also alternatively interpreted epistemically (as theories about the possibility of knowing the world out there) as opposed to ontologically, thus yielding another dimension (Fig. 1; gray axis) corresponding to the dependence arising from the impossibility of directly perceiving reality (philosophical skepticism, solipsism, transcendental idealism). Panpsychism is the view that phenomenal properties are fundamental and ubiquitous constituents of matter (and thus sometimes is contrasted to physicalism; Goff et al. 2017), but without committing to other statements. Finally, dualist doctrines maintain that mental and material objects coexist but are ontologically different by differing on the degree, ranging from strict separation (substance dualism), weak separation (property dualism) and identification (which degenerates semantically into neutral monism). Thus, there is a vast spectrum of doctrines that depict the objects of reality as hybrids resulting from the coalescence of material and mental entities to different degrees (Fig. 1). However—due to the dearth of precisely definable concepts—the boundaries of the scope of these philosophies are fuzzy and ever-changing, so this tentative outline should be regarded with prudence.

Is there any way to arbitrate among this flurry of theories? We can constrain the space of plausible explanations by heeding the few trustworthy clues at our disposal. The essential clue at our immediate reach is, rather prosaically, that the common factor to anything we can know about any world object is *oneself*, i.e. the ever-present observer. This remark is the *observer bias.* It is a bias in the sense that the alternative of perception or knowledge without observer is inconceivable (without dismissing it straightaway). Along with the recognition of the observer bias, I will set forth my case by advocating the agnostic version of philosophical skepticism. When confronted with several competing hypothesis, and in the absence of conclusive

evidence, one should suspend judgment (epoché, in the classical Greek sense of Pyrrhonists and Academic Skeptics), and if forced to choose, pick the option that makes the least assumptions—the simplest one (Occam's razor). The basic rationale for this is that any unwarranted assumption increases the chances of leading to mistaken predictions, so remaining agnostic under uncertainty is safer. The rejection of certain knowledge in favor of probabilities as a guide to behavior was perhaps first avowed by the Academic Skeptic Carneades (Laertius ca. 200/1925). Although this doesn't guarantee to find the right answer, it is the choice most likely to generalize well to other domains, the computationally cheapest, and thus the most efficient from a Bayesian statistical perspective (MacKay 2003). Thus, any indirect or assumption-dependent concepts should be regarded with suspicion. This leads naturally to a view that ascribes fundamental importance to mental events, because these are our only (apparent) primary source of knowledge about the current state of the world.

In the context of phenomenology, the term epoché is also used to refer to "bracketing" or phenomenological reduction, an approach used to (phenomenologically) understand the structure of consciousness stripped away of assumptions about the world (Husserl 1913). It is worth to note that this type of suspension of judgment or epoché is different from the one (classical Greek) propounded above in that prior beliefs that are repressed or put aside in phenomenological reduction are inherent and automatic inferences that affect subjective experiences, whereas classical Greek and Cartesian epoché is used as a way to arbitrate among competing theories—the intended sense here. The focus of phenomenology is on understanding pure consciousness, on the mechanisms underlying the phenomenal observer beliefs (phenomenal selfhood properties), whereas this article focuses on what can be said about the relationship between the functional and phenomenal aspects of the subjective experiences of observers.

We will focus particularly on the subjective aspects of experience (phenomenal experiences[1]) and of knowledge. Although it is an established fact that much of our conscious thinking emanates from unconscious processes (Nisbett and Wilson 1977), and that these (brain) processes actively generate hypothesis about the state of the world (Helmholtz 1860; Barlow 1969), I will ignore mostly and purposefully unconscious processes. The reason is that a discussion of unconscious processes can be distracting for the current purpose, which is not elucidating how our beliefs are actually generated, but how we reason and justify our beliefs and knowledge to ourselves, and what can we learn from this idiosyncratic way of reasoning about the relationship between physics and consciousness (without letting this omission jeopardize logical consistency).

In the following, I will argue that the realization that we suffer from an observer bias and the judicious application of philosophical skepticism can be used advantageously to elucidate why and how we represent the world. Then, I will introduce

[1] I will use the terms subjective experiences, phenomenal experiences, and consciousness indistinctly to refer to the same concept. Specific term choices only reflect emphasis on subjective or objective perspective.

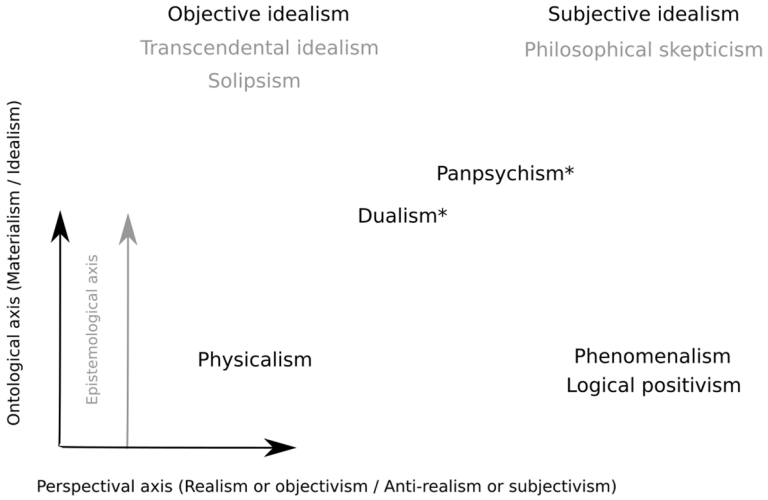

Fig. 1 Diagram representing through proximity relationships a few representative theories in philosophy of mind, grouped by three themes: Ontology or idealism-materialism (black vertical axis); Perspective or subjectivism-objectivism (horizontal axis); and Epistemology or whether knowledge is constrained by ontic or epistemic causes (gray vertical axis). *: panpsychism and dualism lie somewhere in the middle of the ontological axis, but without committing to any side of the perspectival axis

the explanatory gap and expound on the observer and existence biases. Next, I will define more precisely the concept of observers to discuss the subject-object relationship, and apply these insights to a few outstanding problems in philosophy of mind. Finally, I will argue that it is possible to make headway in the hard problem of consciousness.

2 A Skeptical Philosopher Wielding Occam's Razor

In its most primitive form, the observer bias just implies that objects must be perceived in order to exist in an ontological sense—"by themselves". This minimal premise can be used advantageously to glean feasible candidate theories, because then non-perceived objects become beliefs and not objects—or at least beliefs about objects. This entails that our conscious experiences about the current state of the world is built on sensations and inferences about the causes causing these sensations. Although a priori genetically wired generative models about the structure of the world condition the type of perceptual inferences we make—and hence our conscious experiences—here we consider these traits as a factor adding variability to the perceptual machinery of inferential processing across observers. This is because this factor pertains to how world representations differ across observers, but our discussion will focus on what can be said about the relationship between physical properties and phenomenal properties that is common to all observers.

What we usually call knowledge is mostly (probabilistic) subjective beliefs. The very fact that there is a world out there is a belief, as soon as we believe that it has

an independent existence from the observer, or that it evolves according to predetermined (physical) rules. That other persons have sentience and harbor phenomenal experiences like us (or not) is also a belief grounded on theory of mind—the capability to build a system of beliefs about what and how other people think. The current state of the world is a matter of beliefs rather than of knowledge, where these beliefs can take different degrees of uncertainty. Since our only (apparent) primary source of information about the current state of the world is our sensations—as already pointed out by Aristotle and Epicurus in the fourth century BC (Shields 2020)—we are compelled to have beliefs about the actual world if we are to predict the future in order to achieve any particular goal. Any doctrine that assumes attributes of the objects in the world can be anything beyond beliefs can be culled with Occam's razor by cause of being overconfident.

George Berkeley (1710) advocated a radical form of idealism where all world objects are a construction of the mind. In other words, the cause of the world is a mind and material (independent) objects do not exist. His philosophy is summarized sometimes with the aphorism "To be is to be perceived". Although Berkeley's refutation of the independent nature of the world may be unsound, I find this aphorism enlightening. It doesn't assert that objects *cannot* exist without the participation of a perceiving observer, but that unperceived objects are not facts about the actual world, but beliefs.[2] This stance falls under the umbrella of philosophical skepticism, and implies that knowledge (true belief) can be impossible: perceiving an object entails its existence in a phenomenal and epistemological sense, but not perceiving it doesn't entail anything. Depending on what is meant by knowledge, this could be also construed as a fallibilist assertion: the fallibilist's more liberal definition of knowledge encompasses beliefs that are considered certain beyond reasonable doubt. In any case, the bottom line is that both skepticism and fallibilism imply that no belief is guaranteed to reflect truth. What is the difference between how something appears and what it is? From the observer viewpoint, only appearances are accessible, and statements about being (in the sense of existing) are inevitably hypotheses. This also includes the very conjecture that appearances are all there is to objects, or that they are actual appearances concealing an unfathomable being. Thus, strictly speaking, observers can meaningfully use the verb "to be" only in the epistemological sense of having beliefs, as opposed to making statements about the ontology of objects. Therefore, from the viewpoint of an observer, to be perceived is the state closest to being that an object in the actual world can ever achieve. The ontological "to be" degenerates to its epistemological version, and we can't directly perceive objects, but only guess.

For observers, knowledge about the current state of the world is updated phenomenally—phenomenal contents have *ipso facto* a status of truth. In turn, suspension of judgment in combination with the conviction that only one's phenomenal experiences are available *ipso facto* leads to solipsism. Although solipsism is not a falsifiable hypothesis (Popper 1994), neither seem to be falsifiable other doctrines

[2] Thus disavowing a bidirectional implication between perception and being (that Berkeley probably espoused).

incompatible with it such as realism and materialism, since any test for the ontological status of external objects can be conceivably tampered with, for instance by devising arbitrary proxies that meddle with perceptual information (e.g. brain in a vat). At any rate, there is no conceivable way to establish the certainty of knowledge about the external world.[3] Solipsism doesn't make any strong statements about the ontology of the world unlike other idealist philosophies such as subjective idealism, which sets forth specific hypothesis about the non-existence of matter or transcendental idealism, which describes with detail purported mechanisms by which mental processes are filtered and shaped. Because of its minimal foundation, solipsism, with its skeptical (or fallibilist) approach, is less prone to bias our reasoning than other more contrived theories.

3 The Explanatory Gap and Our Biases

We can certainly ascribe to our sensations a certain ontological status, but the existence and properties of an external world remain hypotheses. The urge to minimize our uncertainty about the external world drove humans to painstakingly build—following the scientific method—along centuries a predictive framework that forecasts most of the events that are of practical importance in our lives. This, perhaps after some musings, leads almost immediately to the central problem of philosophy of mind: the explanatory gap. The explanatory gap (Levine 1983)—also called the hard problem of consciousness (Chalmers 1995)—is the problem of explaining in a unified manner the split between the (objective) physical properties of the actual world and our (subjective) experiences: Can third-person quantitative science explain the first-person aspects (phenomenology) of mental states?

Is it possible to bridge this gap? Some philosophers, notably David Chalmers (1996), hold that the gap follows from an ontological difference that eschews any attempt at being bridged by any physicalist theory, that is, any theory that attempts to explain the world in purely physical terms. For others, such as Thomas Nagel (1974), an explanation is beyond our capabilities because currently we lack the required knowledge to understand the question, let alone to provide an answer.[4] Similarly, Levine (1999) argued that the explanatory gap is a symptom of our ignorance of nature rather than a fundamental gap in nature. Others (Dennett 1988; Frankish 2016) dismiss altogether the existence of an explanatory gap as a semantic gimmick or a delusion.

[3] It is interesting to note that in this sense, solipsism is similar to anti-realism (Dummet 1963), i.e. that the "truth" of statements rests on logical validity and soundness, rather than on its correspondence to the state of the world out there. This follows necessarily from the impossibility to cognize or perceive directly the external world.

[4] Nagel (1974) also had the crucial insight that any forthcoming physical explanation is unlikely until we set about understanding better the problem of third versus first person perspectives.

3.1 The Existence of the Explanatory Gap is Not Surprising

Nonetheless, the existence of an explanatory gap is hardly surprising when one contemplates that we are intrinsically wired to appraise the causes of our environment and then act on this appraisal: most of our thoughts and emotions are engaged in the complex task of deciding what to do next. Indeed, we have devised a predictive framework for the world (the natural sciences, including neuroscience) to explain and foresee the causes of our sensations,[5] but not the sensations themselves.[6] This is only natural, since this is all we need to predict any future world event of practical importance. The building of the natural sciences is an extension (into our environment and language) of our innate drive and ability to predict the (especially immediate) future, which in turn inheres in our beliefs about and our drive to apprehend the causes of our sensations and objects of the world. Importantly, this is not to be taken as a dismissal of psychophysics and other scientific disciplines that study our sensations: psychophysics comprises many important stimuli-sensation correspondences, such as the measurement of detection thresholds and the description of sensation intensity as a logarithmic function of some stimuli modalities (Fechner 1860). In fact, psychophysics is perhaps the discipline closest to studying the hard problem of consciousness—but still falling short of it with the currently available methods. This is because its applicability is limited to reportable sensations induced by *external* physical causes—which is precisely what brains were selected to do—so typically it will only uncover the stimuli-sensation relationships pre-determined by our brain machinery, which are strongly biased by natural evolution (cf. Sect. 3.3).

Physical variables such as length, geometry, temperature, velocity, etc. are not only states of the world, but also causes of our sensations. These variables, which we can conjecture to exist independently of minds, evoke correlations among our sensations across different modalities, and therefore are amenable to transformation into estimates of the true hidden causes that determine the actual world (Barlow 1961). Physical theories are hypotheses about the inner workings of the world obtained as the result of thoughtful reflection and careful verification. The belief that physical theories can explain all aspects of the universe is called physicalism. The epistemological value of physical theories lies in that, while not being amenable to

[5] In particular the sensations derived from exteroceptive senses (concerned with causes located outside of our bodies) such as sight or hearing, but also proprioceptive (body position and balance) and interoceptive (visceral and other internal signals closely intertwined with feelings) senses.

[6] One could argue, for example, that colorimetry explains color perception. But there is an important distinction between a predictive and a descriptive framework. Colorimetry describes which properties of the light incident on the eye will result in different perceived colors. This is useful: for example, in photometry, the radiant flux of light at each wavelength is weighed by a luminosity function, which allows to estimate the *perceived* brightness of a stimulus (because human sensitivity to light depends on its wavelength or hue). However, this luminosity function is purely descriptive (and it disregards individual differences): it is based on averaged subjective brightness judgments from a sample of humans. So colorimetry is not predictive in the sense that we cannot use it to explain the specific form of the luminosity function, or more generally the mapping between light radiance and luminance (radiance weighed by the luminosity function), or more importantly what kind of brain manipulation would cause an observer to, e.g. see less colors (colors as hue qualia) or see swapped colors.

direct perception, they afford predictions of the trajectory of the universe to a reasonably practical degree.

In contrast, phenomenal contents such as red, pain, or fear are not beliefs about causes, but our "owned" or *ipso facto* sensations or mental images, readily storable in memory for further examination. They bear no predictive power by themselves, but are directly apprehensible in consciousness (Dennett 1988), unlike abstract physical theories. Subjective experiences are imposed to us and perceived directly rather than inferred; hence their perception can dispense with other mediating variables. In general, it is precisely because we do not need to conjecture additional mediating variables that, in the pragmatical sense of the physical sciences, we already know our phenomenal experiences *ipso facto*, even sensations that are only incomplete depictions of physical properties.

These aspects of conscious experience are called *qualia*. Sometimes the term qualia is used to refer to any phenomenal instance, but usually it refers only to sensory experiences, and in particular to those experiential aspects that are not fully accounted for by physical variables. Frank Jackson (1982) defined qualia as "features of the bodily sensations especially, but also of certain perceptual experiences, which no amount of purely physical information includes." Here, we will more specifically (but still loosely) denote by qualia those aspects of phenomenal experiences that can be swapped, reduced, or eliminated without creating (potential) inconsistencies between our internal representation (brain internal generative model) of physical variables and our afferent sensations (transcendental consistency), and between the different physical variables in our internal model of the world[7] (internal consistency). In other words, the attributes of phenomenal experience for which we cannot find a one–one mapping to physical variables would be qualia.[8] Thus, for example, the quality of "what-it-is-like" seeing red (or fear, or the smell of clove) would be a quale, as opposed to the perception of object lengths,[9] because the former has no one–one relationship to physical variables (and thus can be swapped with, say, blue), whereas the latter does. Notice that under this definition what constitutes qualia is conditioned on our ability to find a relevant mapping to physical variables.

Being baffled by qualia—which are an essential part of our daily lives—can seem odd, especially when one finds the abstract physical theories less baffling. These theories describe the mechanisms of the known universe in an uncannily accurate way. The origin of this paradox lies in that although physical theories are the most

[7] This mapping may vary across subjects. For example, the hue qualia of a color blind subject are less informative than the hue qualia of a standard subject, so their respective mappings to physical light wavelenghts will differ accordingly.

[8] Although given our current state of knowledge any type of conscious experience deserves some explanation, I think that those aspects that have no clear relationship to physical variables are the most baffling but also the most likely to eventually give us an inkling of how to gain ground in the hard problem.

[9] One could argue that the Müller–Lyer and other optical-geometrical illusions suggest that one–one mappings for length perception are also moot. However, optical illusions are rather a manifestation of the difficulty of ascertaining the causes of our sensations, which can render perception ambiguous. But even if we cannot always perceive length accurately, we harbor an internal representation of length which we can consistently and univocally relate (as a concept) to a physical variable.

successful tool at our disposal for predicting the world, they make no reference whatsoever to subjective experiences and in particular to qualia. This is essentially because (by definition) qualia bear no physically relevant information about the state of the world, and therefore have been denied admittance to the natural sciences. But another factor also contributes to the unaccountable status of phenomenal contents. Unlike physical objects or causes, which rest on events taking place in the environment, phenomenal contents originate from neural activity in our brains. Indeed, the only known domain of the actual world known to covariate with phenomenal experiences—from qualia-rich emotional sensations to abstract thoughts such as future planning and mathematical structures—is (some subset of) brain activity, which implies that the rest of the actual world is beyond direct apprehension and virtually clouded by ignorance (Fig. 2). Ironically, unlike with the rest of the actual world, we cannot act upon the physical substrate of phenomenal contents, nor observe it directly, unless we intrude into our skulls or expose and poke our brains, which is to this day an awkward and risky procedure. The inaccessibility of our brains is a major hindrance in the quest to develop a psychophysical theory of consciousness. We can learn a lot about the physical objects of the world that cause our sensations by repeatedly acting and observing the effects of our actions, but not so with the physical objects that cause our consciousness.[10] References to qualia are absent in physical interpretations of the world. Thus, when confronted with difficult questions about the nature of the universe, we tend to think of the universe as ruled by a framework of arcane laws and to dismiss our subjective experiences as fickle and delusive. But naturally this approach is doomed to fail when applied to explain the gap between physical and phenomenal properties.

The existence of the explanatory gap is not surprising; however this is so not because it is a fictitious problem—as Dennett (1991) upholds—but because those aspects of the phenomenology of mental states that have no definite and unequivocal physical counterparts (i.e., no obvious one–one mapping) are precisely those aspects that are ignored by the physical sciences as inessential in the sense of not being elementary physical variables.[11] Although its existence may not be surprising, the gap itself is intuitively bewildering. Just consider that we are comfortable with "explaining" the (unperceivable and hidden) causes of our sensations with probabilistic beliefs grounded on inferences, and usually also with taking for granted our sensations, but when we attempt to explain qualia in the same way as we explain hidden causes we are puzzled over our incompetence. A straightforward, yet not

[10] Electrical brain stimulation studies have been able to induce many different phenomenal experiences. For example, stimulation of a visual area found to be selective for blue-purple, indeed elicited a blue-purple visual percept in human subjects (Murphey et al. 2008). However, these studies are typically limited to finding "hard-wired" brain locus-phenomenal experience mappings which provide little mechanistic insight (cf. Discussion section).

[11] Phenomenal properties that apparently bear null physical information have no place in physics, but at the same time they are the most puzzling properties. At any rate, even if qualia had a straightforward one–one correspondece to physical variables, it's not clear that there would be a way to address this issue.

fully satisfying, reason is that questioning the properties of qualia has no evident benefits for our survival.

3.2 The Observer Bias

Rather than being baffled by our inability to explain first-person experiences with physics, thereby calling into question our only source of knowledge about the current state of the world, a more fruitful approach could be to query the subject of phenomenal experiences: the observer. The formulation of the question "Why are mental states composed of subjective experiences?" is deceptive. First of all, we must realize that although we can doubt the causes of our sensations, we can't doubt our sensations themselves.[12] Since phenomenal experiences are the only foundation on top of which conscious thinking is built, and Occam's razor prompts us to search for the simplest explanation, in time our wary gaze should turn to the next obvious element in phenomenal experiences: the subject. Maybe phenomenal experiences are baffling because we are too focused on their causes, or because we are predisposed to only consider as intelligible those concepts that *I* have inferred, and not those that are surrendered *ipso facto* (at least apparently) without thinking effort. What is more, it seems that there could not be any qualia in the first place without an observer. A better question, thus, may be "why are my mental states composed of subjective experiences?", in the sense of "why do I have subjective experiences?".[13] Omnipresent in all phenomenal experiences is the observer. Although this may sound obvious, emphasizing this notion is an essential ingredient to understanding the explanatory gap because, although it's hard to blame ourselves for it, we tend to take for granted our own existence. Subjective experiences (whatever is their real cause, even in a dream) are the foundation of our beliefs about the current state of the world, but the presence of an observer is a *sine qua non*. And given that all evidence indicates that phenomenal experiences are organic to the texture of the world, applying Occam's razor leads us to questioning the observer.[14] Thus, asking questions about the nature of phenomenal contents is a fact conditioned on the existence of an observer who "owns" those phenomenal contents and is capable and motivated[15] to ask them. This constraint on the ensemble of possible phenomenal contents constitutes the *observer bias* (or *phenomenal observer bias,* for reasons which will become clear in Sect. 4.1). Because the observer bias is partially responsible for the opaqueness of the explanatory gap, bearing it in mind is essential to advance our understanding.

[12] We can incorrectly infer the causes of our sensations, but we cannot incorrectly (phenomenally) sense our sensations. For instance, we can wrongly believe that someone just knocked on the door, but we cannot be mistaken about the sensation itself that we just heard a sudden stroke.

[13] Here we are assuming the existence of I. For those who deny I's existence: in the following, we will refer to I's as *observers,* and Sect. 4.1 will define "observer" and expound on its relevance to consciousness.

[14] A good name for the change of perspective from questioning phenomenal contents to questioning the observer could be *eversion of perspective*, by analogy with sphere eversion (i.e., turning inside out).

[15] Or, in classical parlance—*puto me dubitare, ergo sum*.

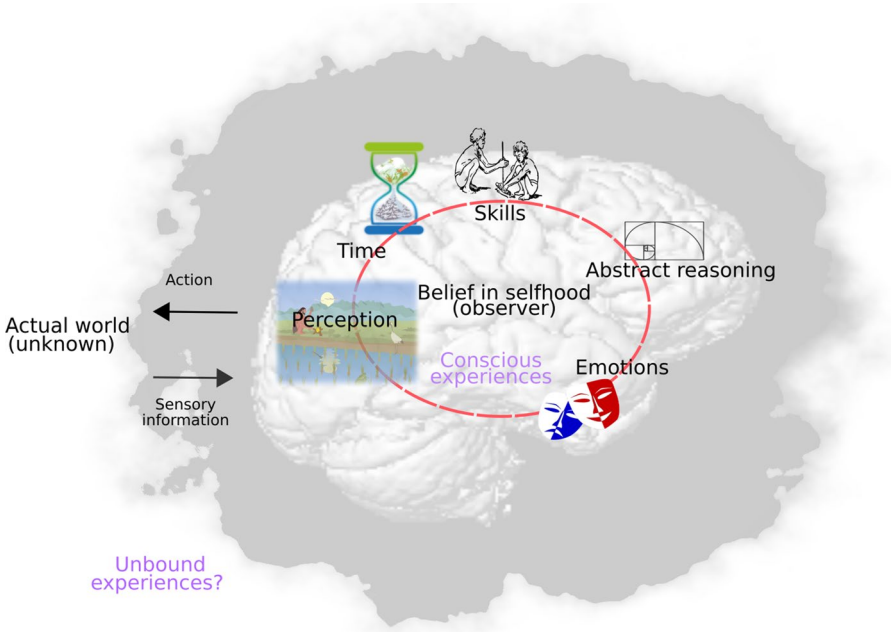

Fig. 2 Schematic of the boundaries of a functional observer defining a region of the world where conscious experiences are "owned" or perceived. All phenomenal contents, whether sensations, emotions or abstract thinking, originate from neural activity in some subset of a brain isolated in a skull, that implements a functional observer. Because the brain is not part of the actual world environment which brains evolved to sense, represent, and act upon interactively (which we still cannot do except sometimes in awkward laboratory conditions), we cannot gain insight into the inner workings of the only substratum generating phenomenal contents. The brain model image was generated with the neuroimaging software SPM (Friston et al. 1995)

3.3 The (Self) Existence Bias

In the same way that one ought to consider the implications that an always-present observer has for the epistemology of phenomenal contents (observer bias), likewise one ought to consider a complementary view: the implications that one's own existing (as an observer) has for the physical attributes of the world (including its inhabiting observers). In particular, this world has to be such that it allows one's own existence. The *(self) existence bias* (or *functional observer bias,* cf. Sect. 4.1) is entailed by the existence of the observer in conjunction with the hypothesis that there is a statistical population of worlds over which different token worlds are to be found.[16] Moreover, we being inhabitants of the world in the current circumstances (such as how long we have been around) in combination with the principle of natural selection (Darwin 1869) has major implications for the kind of creature we can be (Friston et al. 2006). In other words, the actual world must be such that it allows the

[16] This argument is also called the anthropic principle (in cosmology).

existence of observers who question the explanatory gap, and the current state of the world can inform what kind of creatures it can possibly comprise. Finally, the self existence bias in combination with the principle of natural selection and Occam's razor serve as a license to dismiss creationist and other ad hoc doctrines (such as Lebniz's monads or Berkeley's subjective idealism) and to set about to glean the most parsimonious yet informative explanations.

4 The Coupling Between Subjects and Objects

We will now set about exploring the possibility, necessity and contingency modalities[17] that emerge by virtue of how the world is split into the two dovetailing pieces that can constitute it: the subject and the object.

4.1 Observers Exist Because They Believe in Selfhood

The concept of observer (equivalently subject) is central to the comprehension of phenomenal experiences (objects). The reason is that all phenomenal experiences are subjective. Although to this, one (realist philosopher) could object that the objects that appertain to physical theories exist independently of observers,[18] a solipsist could in turn step in and contend that physical theories are just beliefs. Ultimately, all perceived or conceived objects are so in virtue of being perceived or conceived by an observer. Although we have been rather casually discoursing about observers (in the loose meaning of someone who perceives things) without pausing to contemplate what is an observer beyond a terse definition, we cannot carry on in this manner any more.

We will define *functional observer* (or functional subject) as any entity (a proper subsystem embedded in a larger supersystem) that gathers evidence—since it must have the ability to sense the world—and updates its knowledge such that it believes itself physically or functionally to be[19] distinct from the surrounding environment (supersystem), and that on this basis it also perceives[20] and acts upon the world to

[17] In modal logic, these two are collectively denoted as alethic modalities.

[18] Realism as in physics.

[19] In the most general possible sense; for example, by occupying a specific region of space (which is not part of the environment), or effecting on the world, i.e., causing events (which were not caused by the environment).

[20] A boat sailing in open sea equipped with a chip log to measure speed can use dead reckoning to constantly estimate its current position, assuming it knows its initial position. This would be enough for the boat (with its sailors) to have a belief in its current position. Of course, without a navigation system providing geographic coordinates, dead reckoning plots will accumulate errors due to set and drift and become increasingly dissociated from reality. However, a navigation system (say, celestial) would allow the boat to match periodically its position estimates (beliefs) with the actual position (inferred from the stars or sun), in the same way that we update our belief in our state based on our senses. This closed signalling loop would allow the boat to have a belief in its position in the relevant sense of believing in selfhood: it can know that it is estimating *its* position, and not someone else's. An archetype of a device that updates its current beliefs based on a continuous stream of inputs is the Kalman filter (Kalman 1960).

implement and secure[21] this distinction. From here on, we will denote this notion *belief in selfhood,* to emphasize that we refer to the belief in the self-other distinction. For the simple example of a functional observer capable of moving its body to specific locations of the world (navigating), this is just bearing a mechanism that estimates continuously one's own location in the world.[22] Another more complex example is the subset of the brain that implements the self-other distinction as distributed neural circuitry comprising functions such as theory of mind, agency and body ownership. A functional observer loosely conforms to the common concept of observer or subject from a third person perspective: "something that looks alive". However, a functional observer should also have "first-person experiences" in the sense that it holds (exclusive and functional) beliefs about what pertains to it and what pertains to the environment (objects)[23] while it interacts with the world. What does this mean? Ostensibly, a functional observer is the subject of a perceptual and cognitive act. But more revealingly, a functional observer is something defined by its own functional beliefs; in particular, a subject functionally believes that it is different from the rest (of the world). This seems paradoxical because it does not seem that an observer needs to believe in selfhood in order to exist. But it actually does: without a functional belief in a subsystem of the world that is accountable for the properties that define it, there would be no knowledge of how to specify that subsystem[24] (self-specification, Christoff et al. 2011), and no knowledge of its existence, i.e., the observer would not believe in itself (nor anyone else for that matter[25]). And of course, with no concept of selfhood through which a functional observer could attribute (itself) knowledge or beliefs, memories on which to ground a self-concept cannot be created.

A *phenomenal observer* (or subject) is similar to a functional observer, except that instead of functionally or mechanistically, it interprets beliefs and knowledge in the usual phenomenal sense of subjective experiences. In other words, this corresponds to the usual concept of first-person observers, or conscious being, like the reader is now. Crucially, if the reader takes the trouble to reread the the previous paragraph under this phenomenal interpretation (substituting the terms "functional" and "third person" by "phenomenal" and "first person", respectively), it will realize that each and every one of the statements *still* makes sense.

[21] Secure the self-other distinction, i.e. persisting, in the sense of counteracting world events that threaten its survival (Friston et al. 2006). An observer that did not secure this distinction would soon cease to exist in a dissipative world.

[22] A priori, we could try to define an observer without the requirement of believing in selfhood, but this eventually leads to a logical impasse. An observer without belief in selfhood is, loosely speaking, like an unmanned watchtower: nobody is watching anything from it.

[23] Importantly, the boundary (body limit) of the functional observer is arbitrary, as long as the body implements a distinction between self-sensing (proprioception and interoception, cf. Footnote 5) and world-sensing, and the observer is able to preserve its integrity while using it to act upon the world. Thus, the body limit is defined by the extent to which the observer can control and sense the body.

[24] Loosely, if a boat has no way to find its own position, in a sense it is not (or at least cannot act as) a boat anymore.

[25] Being unwittingly seen by someone else does not qualify as existing for someone else in the relevant sense: we are concerned with the beliefs, and not with the body, of the functional observer.

Hence, functionally believing in selfhood enables subsystems not only to define the limits of their body, knowledge, and agency, but at the same time brings about functional states or "experiences". And likewise—substituting "functional" by "phenomenal"—phenomenally believing in selfhood brings about phenomenal experiences or states. In other words, belief in selfhood is a *necessary* condition for the existence of both functional and phenomenal experiences. And this is plausibly so because, in the same way that functional "experiences" are what computing activity appears to be for a functional observer, phenomenal experiences *are* what neural activity *appears to be* to a specific entity believing in selfhood: the phenomenal observer.[26] Indeed, for conscious humans, the functional processes simulating the belief on (or current estimation of) the actual world take place in brain regions that implement functional observers, and it is precisely the activity of these brain regions that we experience as phenomenal contents.[27] But the bottom-line here is that the instantiation of both functional and phenomenal observers necessitates belief in selfhood.

We could also ask whether the existence of a functional observer is a *sufficient* factor to bring about functional states. And likewise for phenomenal observers. In both cases, our experience indicates that as long as the functional (or phenomenal) observer structure is active in a steady regime[28] (i.e., roughly speaking it is constantly assessing the state of the boundary between the self and the other), its existence ought to be sufficient for it to possess functional (or phenomenal) states[29] (at least in the world we know). The reason is that in the relationship between an observer and its *internal* states, subject and object are always coupled: in the same way that a functional observer is implemented or defined by its internal computations, a phenomenal observer is implemented or defined by its phenomenal contents.

As mentioned in the previous section, the self existence bias entails that selection pressures should have afforded creatures a mechanism to monitor their own states (a current state estimate is a current belief in a particular state) in order to foresee the immediate events taking place in the environment and then accordingly act to preserve one's own (or species) integrity. Crucially, functional and phenomenal observers do exactly this. Which leads to the following observation, central to the explanatory gap: *a functional observer is exactly the kind of structure required to realize a phenomenal observer*. Thus, our existence (as phenomenal observers) is strong evidence in favor of the existence of a functional observer, which in turn—in combination with the natural selection principle—is strong evidence in favor of the

[26] This idea strongly resonates with solipsism: being necessitates believing in.

[27] This has been suspected to be so for a long time, but despite our best efforts, we are intrinsically compelled to believe that our phenomenal contents *are* the world out there, even after learning that it's not true (Helmholtz 1860).

[28] Note this differs from the common usage of concepts of specific (observer) persons ("David") or personal pronouns ("you"), since we still consider David to be David even when he is unconsciously sleeping.

[29] This is the same as saying that an observer that ceases to work as such is not an observer any more: observers must be "functioning" ("being") permanently to be considered as such.

existence of a world that is compatible with the presence of self-sustaining entities or life,[30] as opposed to an ad hoc world.

Let us consider the case of a creature who is sophisticated enough to examine the explanatory gap, as we are. Such creature clearly entertains phenomenal contents, a concept of a thinking self, and beliefs about what it knows and does not know. Such a creature must be a phenomenal observer. Crucially, this means that only creatures with attributes of phenomenal observer could find themselves in the state of affairs of the author or the reader, whether it is e.g. pondering over the explanatory gap or the texture of consciousness. This implies that however unlikely the existence of phenomenal observers might seem, we can set this qualm aside on the grounds that the work of a natural selection process laid out precisely this state of affairs, because *it selects* functional observers. Using again Occam's razor, we can also conclude that alternative scenarios like brain in vat or simulated worlds are implausible.

4.2 Are Hollow Observers Observers?

Subjective experiences indicate that (human) phenomenal observers perceive the world in phenomenal states comprising qualia. On the other hand, there is no evidence for phenomenal observers stripped away of qualia (we will examine further this in the next subsection). What about non-human phenomenal observers? It seems plausible that e.g. monkeys too harbor some kind of qualia, even if they can't report it verbally. On the contrary, in the case of simpler forms of life such as viruses (and certainly inorganic objects such as rocks) this seems implausible. Hence, it is conceivable that at some point in evolution (human) functional observers started harboring (or at least believing they harbor) subjective experiences. This would entail that the (contingent) perception of experiences is not a property of the world (objects), but of observers (subjects).[31] This is equivalent to saying that the noumena, or the persistent structures lying "out there", rest on experiences, or at least they consist of entities that subsume experiences.[32] Thus, experiential contingencies depend on the observer, as opposed to on the world. In particular, on whether *there is* an observer.

The concoction of an observer devoid of subjective experiences has been explored under different schemes in philosophy of mind. For example, philosophical zombies (or p-zombies) are fabricated creatures that are prescribed to behave exactly like humans but lack any type of conscious experience or qualia. David Chalmers (1996) most famously and eloquently used p-zombies in an argument in order to refute physicalism.[33] Another similar thought experiment is China Brain (Dneprov 1961; Block 1978), where each member of the Chinese nation perfectly

[30] Abiogenesis and autopoiesis.

[31] This would imply that, if particular instances of qualia underwent changes such as appearing, disappearing or, say, become smeared, this would be just a manifestation of analogous mental processes occurring to the observer.

[32] Note this resonates strongly with panpsychism.

[33] The argument coarsely says that given that the scenario of a world inhabited only by p-zombies seems plausible, it would follow that phenomenal states (consciousness) cannot be explained with physics.

simulates one neuron of the brain. All these schemes fall under the umbrella of the absent qualia hypothesis, namely that a system that functionally duplicates the mental states of a normal human being could lack phenomenal consciousness and thus qualia (Tye 2006). Although the existence of such phenomenally hollow systems or creatures seems possible in some regards (Chalmers 1996), without further premises it is unclear to what extent this admission is warranted. At first approximation, a skeptical view would suggest that an observer who totally lacks experience is degenerate to the extent that it is not really an observer anymore, i.e., a trivial[34] or hollow observer.

4.3 Is Unbound Consciousness Consciousness?

We can also flip the arrow of the subject-object relationship (or everse our natural egocentric perspective) by asking: is the existence of qualia (or generally consciousness) without observer possible?[35] The case of unperceived qualia is illustrated by a popular thought experiment commonly associated to Berkeley: does a falling tree in an uninhabited forest[36] make a sound? This device is often used to challenge our intuitions about the ontological status of unperceived objects. In other words, it asks whether phenomenal experiences can be unbound from observers. Although it's not clear that the concept of unperceived (unbound) qualia or experiences is even consistent, the concurrence of subjects and (phenomenal) objects cannot be untangled in our actual world, which is precisely what engenders the observer bias (see Sect. 3.2). Thus, there are no reliable clues to venture an answer generalizable beyond our actual world. But it's worth noting that, because of the symmetry of this logical conundrum, the existence of p-zombies is plausible in a way similar to how there could be (unheard) sound in uninhabited forests.

When we refer to subjective experiences we often omit that they always concur with an observer because they are defined by pertaining to an observer. Thus, we could either enlarge the semantic bag of the phrase-concept "phenomenal experiences" to incorporate observer-independent consciousness (unbound consciousness), in which case we are compelled to accept panpsychism and therefore myriad of forms of consciousness (e.g., rock-like psyche) unlike ours; or maintain our conventional definition of phenomenal experiences as *subjective* experiences.[37] We will choose the latter for simplicity, and because unbound experiences are unlike any conscious experience we would ever be able to have[38] (Fig. 2), and arguably even harder to imagine than what it is like to be a bat (cf. Nagel 1974), since bats at

[34] In the mathematical sense of the word denoting the simplest possible structure.

[35] Note this is not the same as observers without qualia.

[36] Interestingly, this was precisely the state of much of the Earth continental surface during the Devonian period, with extensive land colonization by plants and the absence of large animals, except perhaps some arthropods (Fortey 2011). The question would ensue then whether we could consider these arthropods as observers entertaining sound qualia.

[37] In which case we should be careful saying things like "what is it like to be" unbound experiences, since there is no *to be like* for unbound experiences.

[38] Indeed, I use the word experience in the phrase "unbound experience" only very reluctantly.

least *are* plausibly observers in their own way. Unbound consciousness *cannot* be perceived: there is no structure to tell apart a subject from the (unbound consciousness) object, there is no sense of subjective perspective, of subject, or understanding for that matter, since perception is a process requiring actively constructing hypothesis (Helmholtz 1860; Gregory 1980). Indeed, for this very same reason an observer never ceases to observe from a first-person perspective. It could be argued that this doesn't really answer the question, which would be fair. The real answer is that currently we cannot know what it even means as long as we don't redefine terminology. However, our choice is consistent with the solipsist view that unperceived objects, as long as unperceived, are nothing beyond beliefs.[39]

Finally, the answer to the next question is key to bridging the physics-consciousness explanatory gap[40]: what is the relationship between functional observers and phenomenal observers? The main point of this section has been that a phenomenal observer necessitates a functional observer as a physical foundation. But is the converse also true? Specifically, are functional and observers two facets of the same entity? Given the arguments put forward in this section, there seems to be no evidence for the contrary, so applying Occam's razor, the answer would be yes. Nonetheless, more rigorously, the answer should depend on the metaphysical possibility of associating different phenomenal events to physical events in arbitrary worlds. We will explore this issue in the context of p-zombies in the next section.

5 Screening a Zombie World Through the Sieve of Skepticism

Now that we have established a philosophical framework (Sect. 2), and are equipped with suitable tools for reasoning (Sects. 3 and 4), we are in position to take on p-zombies. In particular, is it feasible to kill p-zombies with Occam's razor? That is, can we rule out the possibility of the existence of p-zombies (hollow observers) by only assuming a skeptical and minimalist approach?

5.1 Behavioral p-Zombies and the Chinese Room

The conceivability of p-zombies is a vexing quandary. P-zombies are inhabitants of a physically indistinguishable world lacking phenomenal experiences. The argument goes that the metaphysical possibility of such world implies that physical truths are not metaphysically coupled to phenomenal truths (Chalmers 1996). At first sight, it seems that p-zombies are conceivable as inhabitants of a hypothetical (physical) world. We can indeed imagine a copy of our world and decree by the sole power of our imagination that it be inhabited by p-zombies. We can even picture the p-zombies roaming and reacting to the environment as we would do and likewise imagine

[39] A last thought case would be the possibility of no qualia and no observer. Although we cannot have any evidence for this, it certainly seems uncannily plausible as a degenerate case.

[40] An alternative name could be the noumena-phenomena explanatory gap.

that they lack subjective experiences. This is easy to do when imagined from a third person perspective.

If the p-zombie is just a behavioral p-zombie (i.e., it is constrained only by behaving exactly like a human), it is not difficult to accept its plausibility. We can simply imagine an exact copy of our world where the creatures that we take to be humans are actually behavioral p-zombies, copycats. For example, by populating the world with sophisticated hollow robots, or by placing our brain in a vat and simulating the world (Harman 1973). A p-zombie that is outwardly like us is conceivable because only mimicking is required from it.[41] However, the conceivability of behavioral p-zombies proves nothing by the very same reason that it only requires mimicking to be instantiated.

This scenario is analogous to the Chinese Room, a thought experiment by John Searle (1980). The Chinese Room was devised to show that a computer executing a sequence of operations (a program) cannot be said to have consciousness or even understanding, regardless of how intelligently it may *seem* to behave. The Chinese Room takes Chinese characters as input and following some algorithm it outputs other characters that to any Chinese speaker resemble human answers, that is, it passes the Turing test in Chinese. Searle then indicates that anyone could sit in the Chinese Room and process the characters according to a set of instructions (algorithm) and likewise output characters resembling human answers without having no understanding of Chinese whatsoever. Therefore, Searle and others (Hawkins 2004) argue that nothing in the Chinese Room understands Chinese, but it merely simulates understanding (and consciousness).

This dilemma is compelled by the difference between explicit and implicit knowledge. The Chinese Room has an explicit knowledge of Chinese in the form of instructions prescribing explicit input–output patterns, readily available through a simple search or application of a rule like a hash function. However, a human speaker has knowledge mostly in implicit form: he/she doesn't have direct access to all answers, but constructs answers off the top of his/her head. A human speaker has no access to a dictionary that maps keys (questions) to values (answers), but instead combines retrieved memories with abstract constructs to narrow down, infer, and select an answer. Although much of these computations happen unconsciously, the human speaker is left with a feeling of agency. Ultimately, the only functional difference between the Chinese Room and a human speaker is reduced to whether the answers are found through a simple rule on a big database (dictionary lookup) or through a complex rule on a smaller database (inference engine). Thus, the only substantial reason behind the stated inability to have consciousness or understanding of machine-like sequences of operations is not in the way (mental) computations are performed, but the (phenomenal) feeling of agency that characterizes human speakers. Hence, it is conceivable that the Chinese Room could have consciousness and understanding by endowing it with an additional mechanism that furnished it with a sense of agency and selfness. The Chinese Room is analogous to a p-zombie: one only needs to substitute the human-like behavior and appearance of p-zombies

[41] Indeed, from a solipsist point of view the real world and the behavioral zombie world are identical.

with the output characters and the algorithmic heart of the Chinese Room with the p-zombie (hollow) shell, respectively.

5.2 Neurological p-Zombies and Self-awareness

Neurological p-zombies are endowed with exact physical brain and body copies. Is their existence still plausible? From a third-person perspective it seems to be the case again. How do p-zombies talk about subjective experiences? They do not have any, but they entertain the functional concepts that, in humans, are accompanied by experiences. Even during covert thinking, p-zombies reproduce our functional neural patterns of activity. These are internal states of the brain that represent diverse facets of the world, so they can be linked to beliefs about physical objects and to particular qualia, like red and cold. This functional preservation ensures the behavioral integrity of p-zombies.

However, the same scenario from a first-person perspective is rather different. The reader can imagine for example his own p-zombie doppelganger. Is this first-person p-zombie conscious? By the very definition of p-zombie, it is not. It has been argued that the p-zombie concept is incoherent (Dennett 1999) because they claim to be conscious, as we do. When we engage in introspection, we can attend to and manipulate mental percepts or concepts that are accessible only in first-person, close our eyes, or evoke mental images. But how can p-zombies do this? P-zombies entertain and manipulate analogous internal representations as well, although without "seeing" them. But this "seeing" is for us equivalent to having conscious experiences, and conscious experiencing is the *only* way we have to manipulate and be aware of such percepts. Is it then plausible to say that a being that claims to be conscious while being devoid of conscious experiences is aware of anything, let alone of itself? This jarring feeling about the shift from third to first person perspective is rooted on the assumption that the brain is solely comprised of matter, and yet physically unaccountable phenomenal experiences arise from it. But actually the brain is comprised of poorly comprehended (physical) objects, which ultimately are part of our beliefs about the (purported) actual world. Thus, this jarring feeling is a product of our unjustifiably strong beliefs in our model of the world.

Although we have been discussing only the the superficial aspects of p-zombies, a fair amount of neural processing happens without the involvement of consciousness, which we need to account for. We can pretend that p-zombies are like us except for lacking conscious neural computations, but we can't dodge the contradiction rooted in their phenomenal hollowness. When prompted about their experiences, p-zombies will explain how their internal representations "look" to them, and just as us they will be unable to describe ineffable qualia. They will "demonstrate" by their actions to be aware of their internal representations, and to perceive something that we are forbidden to call qualia, but which exists as patterns of neural activity. Thus, although in the actual world we could say that subjective experiences have a physical causal role, in a zombie world (a world where phenomenal experiences do not exist) the causal equivalent would be the functional associate of those experiences,

i.e., plain neural activity. Although p-zombies have no sentience, "they" can be set in motion ("motivated") by their patterns ("thoughts").

This suggests that we might be conflating having experiences with belief in selfhood.[42] The only difference between p-zombies and us is that we define them as lacking subjective experiences. But when we imagine p-zombies, we don't envisage a crippled observer, but rather a degenerate form closer to not being an observer (a hollow observer, cf. Sect. 4.2). The real cause of the dilemma might not be whether they entertain subjective experiences, but whether they entertain a belief in a self (that can claim to entertain such experiences), or a belief in selfhood (as defined in Sect. 4.1). We cannot assign unbound consciousness, let alone a full blown self concept, to a chunk of matter such as a rock because it is not a functional observer (see Sect. 4.3). The substantive conundrum implied by saying that p-zombies lack sentience or qualia, is that they are creatures equipped with a machinery capable of outwardly displaying self-awareness, while lacking self-awareness. This can be exemplified with the following twist: if we were asking whether rocks in a zombie world are conceivable, we would probably say yes. But the reason is not only that we do not attribute consciousness to rocks, but more meaningfully because we do not attribute self-awareness to them. Indeed, in a world where there exist creatures that entertain phenomenal experiences, the question of whether rocks entertain subjective experiences would be equivalent to the question of whether they have associated epiphenomenal experiences while not being functional observers.[43] Hence, p-zombie-rocks are arguably not baffling because we intuitively recognize that they are not functional observers and thus lack belief in selfhood. Contrastingly, (neurological) p-zombies are endowed with functional observers that enable and *enforce* their belief in selfhood (see Sect. 4.1). In other words, p-zombies by definition believe in selfhood, but at the same time are phenomenally unaware of themselves.

5.3 Philosophical Zombies are Contrived to Believe in Themselves Functionally but not Phenomenally

P-zombies are baffling because they believe in selfhood and therefore they could have subjective experiences, yet (by definition) they don't. From a first person perspective, imposing a zombie world would take away all your phenomenal experiences, which amounts to everything a phenomenal observer is. But p-zombies also functionally believe to be conscious, and we know they are not lying or feigning to be conscious, because they are functionally identical to us, and we would as well claim to be conscious. Confusingly, this propositional attitude is a hallmark of consciousness, because (Sect. 4.1) a phenomenal observer believing in selfhood is conscious. P-zombies say "I", demonstrating a concept of self, but a p-zombie-rock would not say "I", since it is not a functional observer. Therefore, p-zombies are functionally equipped to have (and act as if they had) phenomenal experiences, but

[42] That is, assuming these concepts could be dissociated.

[43] This would entail panpsychism.

they do not, whereas p-rock-zombies are unconditionally incapable to entertain subjective experiences.

The p-zombie riddle arises likely from an essential point that is often overlooked: when we impose the condition of lacking consciousness, we are ultimately imposing the condition of being unaware of having consciousness. This is because a functional observer is necessary for consciousness (Sect. 4.1). Functional observer-independent (unbound) conscious experiences should not be called conscious experiences, but something else to avoid confusion, say, unbound consciousness[44](Sect. 4.3). Thus, when we impose on p-zombies endowed with functional beliefs the property of not being conscious, we are imposing a condition that induces a state of affairs for which we have absolutely no evidence: (1) by referring to p-zombies as "they", we implicitly assume they are (hollow) functional observers; but we have no evidence that such entity (with absent qualia) can exist (Sect. 4.2); (2) even if we somehow accepted that p-zombies were functional observers, there is no evidence that the dissociation between functional and phenomenal beliefs assumed by p-zombies is possible. From a skeptical stance, there is no way to support these notions.

The answer to the next statement would solve the p-zombie dilemma: what attributes of a particular world would automatically make any functional observer also a phenomenal observer? Evidence indicates that they are identical in the actual world (Sect. 4), but we cannot answer this question in general. Is there a way to find evidence against p-zombies, rather than absence of evidence? Daniel Dennett took up the idea of p-zombie beliefs further forcefully by conceiving *zimboes*—p-zombies that have second-order beliefs—to argue that the concept of p-zombie is self-refuting: "Zimboes think^{Z} they are conscious, think^{Z} they have qualia, think^{Z} they suffer pains—they are just 'wrong' (according to this lamentable tradition), in ways that neither they nor we could ever discover!" (Dennett 1995). Zimboes have not only functional beliefs, but also some phenomenal beliefs (second order beliefs). Thus, the concept of a zimboe (a human without phenomenal experiences but nevertheless with second order beliefs) pushes forcefully its conceivability to the limit of consistency, and the task of dismissing zimboes conceivability is almost trivial. The point in question is still whether something can be said about the conceivability of conventional p-zombies without endowing them with new properties.

5.4 In a Zombie World, Who Would Hunt Out p-Zombies?

Setting aside the theoretical aspect of the dilemma for a moment, let us assume that we wanted to test empirically whether p-zombies exist. How would we go about detecting p-zombies? Clearly we could not, since they are functionally identical to us by definition, and phenomenal experiences are private. This raises the question of whether the coexistence of p-zombies with phenomenal observers is metaphysically consistent.

[44] Borrowing Kant's terminology (1787/1998), we could even call them noumena if we assumed a panpsychist view.

If we make the simple assumption that each world is nomologically uniform within its state space (i.e., physical theories apply the same to all its constituent objects), it would be impossible for a phenomenal observers to coexist with p-zombies. This is because as soon as, in a particular world, there exists one functional observer who is also a phenomenal observer, automatically all functional observers should also bear phenomenal contents. This would imply that zombie worlds *cannot be phenomenally observed*, that is, no conscious creature could ever see a zombie world. Therefore, even if there was a way for phenomenal observers to detect p-zombies, zombie worlds would be still unobservable. Ultimately, if all creatures are p-zombies, there are no phenomenal observers to give account of this fact, but if there is at least one phenomenal observer, then there would be no p-zombies to be hunted out with Occam's razor even if a method to detect them was available. Thus, the closest that phenomenal observers could get to a zombie world is to believe in it, since they cannot ever find themselves in one.

5.5 Types of Possibility

P-zombies are equipped to believe in selfhood, but they don't have phenomenal experiences. Under which conditions could this be possible? Chalmers (1996) argued that "we can conceive of a world physically indistinguishable from our world but in which there is no consciousness", and that we can infer the metaphysical possibility[45] of p-zombies from the conceivability of p-zombies, since for phenomenal concepts, conceivability (logical possibility) implies (metaphysical) possibility (Kripke 1980). Thus, the metaphysical possibility of a physically indistinguishable world with some p-zombies would imply that physical truths do not metaphysically necessitate phenomenal truths.[46] This was an attempt to impugn the physicalist claim that phenomenal contents can be explained in physical terms, in the sense that physical theories have no bearing on phenomenology (Sect. 3.1). Certainly we can conceive a zombie world under suitably weak constraints, but this can always be done by defining loosely or unloading semantically the term "physical", since in the definition of this term lies a sizeable part of the p-zombie quandary.[47] The conventional definition of physicalism is that "a property is physical if and only if it either is the sort of property that physical theory tells us about or else is a property which metaphysically (or logically) supervenes on the sort of property that physical theory tells us about" (Stoljar 2017). Thus the definition of physicalism hinges on the state of the art in physics.[48] This demotes the importance of answering the question of whether physicalism is true in our world, because whatever can be said

[45] That is, physically possible in other (hypothetical) world.

[46] Alternatively, the existence of inverted qualia (a thought experiment where instead of lacking qualia observers perceive different types of qualia while reacting identically to objects; Locke 1689) would also allow to reach the same conclusion.

[47] Just consider that some physicalist philosophers dismiss the dilemma straightaway as self-refuting by arguing that consciousness is explained only by physical laws, and thus a zombie world would be impossible.

[48] The question of how to choose a representative theory of physics is called Hempel's dilemma.

of physicalism is tied to the predictive power of a conjectural theory ever under construction. Therefore, we are not in position to affirm whether physicalism is true or not for the same reason that we cannot claim that our physical theories are exactly accurate (our theories are ultimately beliefs).

At any rate, the question of the metaphysical possibility of p-zombies (let this proposition be $\Diamond Z$)[49] remains open, since (meta)physically there might be restrictions on what universe configurations could possibly exist. A relevant question here is whether there are metaphysically necessary and metaphysically possible[50] properties to world objects (Kripke 1980). An example would be the fact that heat is decoherent molecular motion: although this was discovered empirically, it is a necessity (empirically necessary; Kripke 1980) in the actual world. Similarly, we have established (Sect. 4.1) that it is necessary for phenomenal and functional observers to believe in selfhood ($\Box S$); and that phenomenal contents necessitate the presence of a phenomenal observer ($\Box O$), because we cannot conceive any (meta)physical situation where these propositions do not hold.[51] In general, necessities are induced by (the belief in) underlying causal relationships afforded by physical or logical arguments that link a priori non-identical semantic concepts. To what extent do these modal properties generalize beyond the actual world?

If we assume that the actual world w^* is nomologically uniform (Sect. 4), then it is false that p-zombies exist in w^* (Z is not true at w^*, i.e. $w^* \nvDash Z$[52]). However, we cannot verify whether others have experiences; we can at most estimate the likelihood of different scenarios (perhaps by resorting to theory of mind).[53] It is unclear whether other worlds could allow Z, so $w^* \vDash \Diamond Z$ remains an open question. Given a set of beliefs about w^*, how would these beliefs map onto the different alethic modalities? We know that p-zombies are not necessary ($\neg\Box Z$), but we ignore what conditions (what type of world) should hold for p-zombies to be possible ($\Diamond Z$) or impossible ($\neg\Diamond Z$). Different scenarios can be nested according to their degree of possibility: the most restrictive scenario would correspond to physical (actual world) possibility (w^*), with metaphysical (m) and logical possibility (a) being more encompassing (Fig. 3). Plain conceivability is equivalent to logical possibility, so p-zombies are logically possible[54] ($a \vDash \Diamond Z$), but to ascertain whether p-zombies are

[49] We follow the convention for unary modal operators: $\Box$ denotes "it is necessary that", and $\Diamond$ "it is possible that". For example, if Z denotes "p-zombies exist", then $\Diamond Z$ denotes "the existence of p-zombies is possible".

[50] For both necessity and possibility, in the sense of being less restrictive than the physical laws of the actual world, but more than arbitrary formal systems that might not be suitable as "physical engines", i.e. as sets of rules and objects that describe (meta)physical worlds.

[51] In Kantian terminology, these propositions would be designated synthetic a priori (Kant 1787/1998), i.e., they were derived based on a non-empirical (metaphysical) basis.

[52] In the semantics notation of modal logic, the symbol $\vDash$ denotes semantic consequence, and can be read as "entails"; $\nvDash$ is its negation.

[53] Hence the skeptical approach we have adopted throughout.

[54] Since we ignore the fundamental mechanisms linking physical and phenomenal events, we lack the belief of the (hypothetical) premises required to declare the proposition impossible.

metaphysically[55] possible ($m \vDash Z?$), we would need to understand the laws ruling jointly the physical and phenomenal aspects of matter, and from there infer under which conditions (worlds) they are possible. Yet with the current state of knowledge, any substantial modal claims lack a basis, and it is unclear whether this could ever be a feasible enterprise.[56] Thus our p-zombie hunt in metaphysical worlds ends in aporia. Another approach would be undermining the logical foundations of $\Diamond Z$, which requires explaining what are hollow observers (Sect. 4.2) more conclusively. Yet, this stumbles upon a similar semantic problem: the hollow observer is a concocted notion for which no evidence exists to say anything meaningful about it. Both in the case of phenomenal or functional observers, the statement "a hollow observer is an observer" is a undecidable statement: declaring it to be consistent or inconsistent depends solely on how we define "observer" by either allowing or not for observers to be degenerate in the sense of empty from contents. But if we allowed the observer to be degenerate, the definition of observer would become virtually useless.

6 Discussion and Conclusion

Contemplating p-zombies is a frustrating undertaking because p-zombies are forcefully defined to incorporate seemingly incompatible properties that we are neither able to dismiss nor to justify reasonably: a functional observer uncoupled from a phenomenal observer. But precisely because of this incompatibility, the p-zombie conceivability dilemma strikes at the core of the explanatory gap. No evidence indicates that p-zombies exist, nor there is support for modal claims of their metaphysical possibility. Although this is insufficient to establish their nomological impossibility, any explanation which accommodates p-zombies must resort to contrived maneuvers that involve stretching semantics or making artificial assumptions. However, we can strike a balance between explanatory power and simplicity by adopting a skeptical approach and using Occam's razor to cull unduly contrived schemes.

We examined the relationship between functional and phenomenal observers, and established that functional observers are the simplest entities that could possibly bring about phenomenal observers. Although a combination of semantic indeterminacy and insufficient understanding of the metaphysical constraints relevant to the explanatory gap precludes finding an conclusive answer, this leads to the following argument about the ontology of the actual world. Given that the existence, characteristics, and origin of functional observers are well substantiated in the context of the actual world's physical laws, that phenomenal observers exist, and that the former is exactly the kind of structure that could engender the latter (in functional terms), the simplest doctrine that accounts for the functional and phenomenal observer coupling

[55] This metaphysical possibility is referred to as "nomological necessity": a statement is nomologically necessary if it is true at all possible worlds ruled by the same physical laws that rule the actual world w^*.
[56] The most we can do is to abduct $\neg\Diamond Z$ from $\Box O$, because if p-zombies were impossible, then phenomenal experiences would necessitate phenomenal observers ($\neg\Diamond Z \rightarrow \Box O$), but abduction is not a valid logical inference.

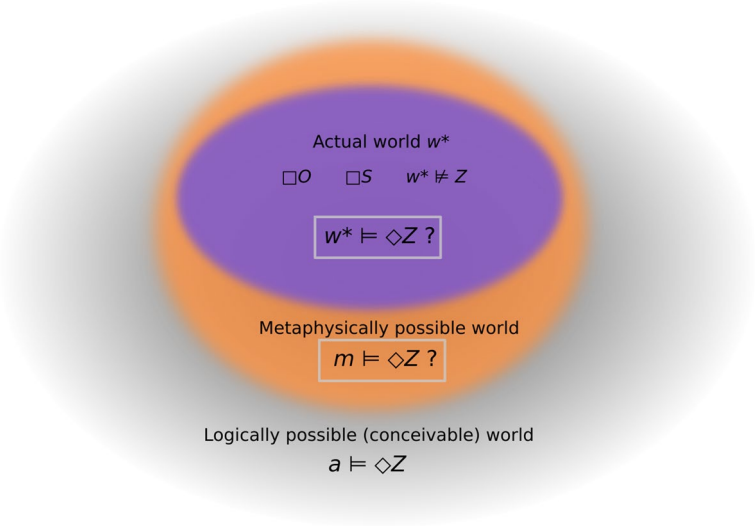

Fig. 3 Depiction of different world scenarios nested according to the degree of possibility of harboring a zombie world: the innermost blob is the most restrictive, corresponding to the actual world, the middle one to a more permissive metaphysically possible word, and the outer to a world only restricted by logical possibility. Whether a zombie world is physically ($w^*\models\Diamond Z$) or metaphysicaly ($m\models\Diamond Z$) possible cannot be answered with our current knowledge

would be one that regards phenomenal properties to be an ubiquitous property of matter. Panpsychism is the doctrine that most closely matches this condition.[57] In other words, *if panpsychism*[58] *were true, then the way we see the actual would make sense with no further assumptions.* Concretely, a functional observer is analogous to a "membrane" between an interior that believes in selfhood and the exterior (i.e. a mechanism to recognize and monitor the boundary between the interior or self and the exterior or other; see Sect. 4.1). Crucially, this is necessary to instantiate the phenomenal observer or *I* (Sect. 4). Further, the functional observer is also analogous to a "world mirror": its inside is a reflection or simulation of the environment *and* of its functional concept of itself or subject[59] (Sect. 4.1). In other words, the functional observer is a (imperfect, but good enough) simulation of the world, held in a physical subspace of the world (e.g. brain), that includes a subject instantiated

[57] Property dualism and neutral monism also would conform, depending on the details. It is difficult to tell whether the distinction between the physical and mental aspects of property dualism is epistemic or ontic, but applying Occam's razor, I am more inclined toward epistemic, since there is no need to postulate two substances.

[58] Alternatively, panprotopsychism (Chalmers 2015), a variation of panpsychism where instead of phenomenal properties, "protophenomenal properties" are posited as the fundamental constituents of the world. These properties are characterized by combining to bring about all forms of consciousness, and by the condition that the understanding of protophenomenal properties enables the complete characterization of phenomenal properties.

[59] Note that this excludes the physical substrate of the functional observer itself (the brain).

through a device that recognizes and monitors its "membrane" or boundary with the environment. In brief, a functional observer defines a physical subset of the universe that encloses both a miniature representation of the environment and a constructed representation of itself that are related to each other through a subject-object relationship: metaphorically, it is a place of the world that is both simulating a theater piece (the world, the object) and *also* simulating an observer defined by its belief of being different from the environment (the subject defined by its belief in selfhood). This is relevant to the ontological status of the world because, together with solipsism and the realization of our own existence, it suggests that the act of having phenomenal experiences might be limited not by the putative phenomenal properties of the fabric of the world, but by the presence of subjects: a functional observer does not open a window to a subset of the world that would otherwise be hidden or lacking phenomenal contents, but rather it creates a subject that can attribute to itself having phenomenal contents. Impaired awareness of one's own internal states is a condition common to many cognitive dimensions, such as body and limb motor control (Blakemore et al. 2002), lack of insight about one's own illness in psychiatric disorders (e.g. schizophrenia; David et al. 2012), and perceptual decisions (Fleming and Dolan 2012). All these manifestations are metacognitive failures underpinned by the misspecification of internal representations or their mismatch with the state of the environment, which goes along with the thesis that an operative functional observer, continuously updating its boundaries with the environment (belief in selfhood, Sect. 4.1), is essential to consciousness.

Under this view, panpsychism is not an endorsement that everything in the world is self-conscious, but the view that the correspondence between physical and phenomenal properties is nomologically uniform all over the world, and that the absence or presence of phenomenal experiences is only determined by the instantiation (presence) of functional observers or subjects—this is reminiscent of panprotopsychism (see footnote 58; Chalmers 2015). Thus, only those entities endowed with functional observers will be capable of having conscious experiences. Loosely speaking, "consciousness" is a ubiquitous aspect of physical objects, but for it to be "seen", its must be "owned" by "eyes" or functional observers, which are akin to "mirrors" where a piece of the universe "sees" itself by simulating both a "seer" and a "seen".

The psychophysical correspondence between phenomenal states and functional states in observers is still mostly uncharted territory that holds the key to the explanatory gap. Is it possible to work out this conundrum empirically? Neuroscientists have been trying to pinpoint for some time the hypothetical minimal set of neural processes that are necessary for consciousness, i.e. the neural correlates of consciousness (Crick and Koch 1990; Lamme 2006). Some relevant neuroscientific approaches are (1) assessing correlations between sensory input, neural activity, behavior, and judgments about phenomenal experiences; and (2) assessing the causal effects of brain stimulation—which has yielded fascinating empirical insights into the localization of many phenomenal experiences in the brain (Selimbeyoglu and Parvizi 2010)—pharmacological interventions, or lesions on phenomenal

experiences, such as eliciting light perception (without using the eyes[60]) through electrical stimulation of the visual cortex (Brindley and Lewin 1968). Some other research topics to the purpose are dreaming, altered states of consciousness, and also the (surprisingly common) condition of aphantasia, where one cannot voluntarily visualize mental images (Galton 1880; Zeman et al. 2015).[61] The importance of understanding the neural footing of internal representations is also becoming increasingly appreciated in philosophy of mind[62] (Jackson 2007). All these approaches are useful expedients to investigate how and what information is accessed by functional observers, and indeed several popular theories have been put forward and are widely discussed (e.g., Global Workspace: Baars 1988; 1999; Dehaene et al. 2003; Higher-order theories: Lau and Rosenthal 2011). Although these investigations provide new knowledge about functional observer configurations that manifest as phenomenal experiences, they have not provided yet new knowledge with regard to the explanatory gap. Brain stimulation research is and will be crucial to the study of consciousness, perhaps as much as to become the only way to gain ground in the hard problem. But our current methods do not allow to go much further than describing correspondences between stimulation intensity and loci on one side, and sensation or behavior on the other side. This is a necessary and fascinating endeavor, but we are still left with the question of why the discovered mappings are such, and how they are induced by the anatomy and function of neural circuits.

Our current inability to easily access the primary physical substrate of phenomenal contents (the brain) in a controlled manner at the mesoscopic level of neural circuits is a major handicap to work out the physics-consciousness explanatory gap (Sect. 3.1). Normal functional observers have been naturally selected to solve physical[63] problems about events happening outside the brain, so by design any functional or (potential) phenomenal contents irrelevant to explaining physical world events are tucked away under the rug of consciousness. Crucially, this includes any intermediate processes located in the brain: any information which is not related to solving high-level cognitive problems about the actual world is shielded from the functional observer in the form of unconscious processes, and this precludes gaining insight into the physics-consciousness relationship. This would require interfering with brain activity in a more precisely targeted manner (at least at the mesoscopic neuronal circuit level) than with the current causal methods of intervention, mostly restricted to occasional intracranial recordings or stimulation during epilepsy surgery, which precludes unrestricted and thorough manipulation. But even tinkering with the brain and observing effects on phenomenal experiences, might still not be enough to move forward in the hard problem of consciousness. We would need to

[60] These visions are called phosphenes.

[61] This is so to the point where aphantasics may not realize that other people can voluntarily evoke mental images.

[62] Since Frank Jackson (2007) came around, he believes that phenomenal contents are internal brain representations that can be accounted for in physicalist terms.

[63] Physical in the broad sense of anything abiding in the environment we evolved to deal with, including other living beings.

find non-trivial links between functional and phenomenal properties, and for this it is expedient to focus on qualia (as defined in Sect. 3.1, and as opposed to non-qualia) phenomenal contents. This is because non-qualia experiences are information-maximizing representations (Barlow 1961) about physical causes of the actual world—which is manifested in their one–one mapping to physical quantities[64]—so they are less likely to afford unexpected findings; in contrast, physical quantity-qualia mappings are often many-to-one (temperature versus hot/cold) or arbitrary or undefined (most emotions). For example, one can imagine a neuroscientist tracking how an object's velocity is encoded along the pathways from the retina to the parietal cortex, and establish the dynamical activity patterns associated to the object's velocity phenomenal percept, thus finding a neural correlate of velocity that contains the same amount of information (perhaps with redundancy for the sake of robustness) as the physical variable—this is an instance of an "easy" problem (Chalmers 1995). However, establishing an association between different hue qualia and patterns of neural activity might constitute progress into the hard problem of consciousness (if for example, this association reveals how portions of the light spectrum can be assigned different mappings to some partition of the range of hues). Hence, direct, causal, mesoscopic scale-precise (yet distributed at the macroscopic level) interventions in the brain may be the only plausible approach to eventually make forays into the physics-consciousness explanatory gap.

In conclusion, I argue that a combination of philosophical skepticism and a preference for simplicity leads to the view that since phenomenal experiences are the sole way to access *ipso facto* reality, (1) the events of the actual world can only be at most beliefs (solipsism) as the most reasonable epistemological doctrine; and (2) phenomenal properties are a fundamental and omnipresent constituent of the actual world (panpsychism or neutral monism) as the most plausible ontological doctrine. Finally, in view to gain empirical ground in the psychophysics of consciousness, assessing the effects of targeted causal interventions in the substrate of phenomenal experiences has the potential to uncover non-trivial relationships between qualia states and functional states.

Acknowledgements I thank anonymous reviewers for improving this article through their thoughtful and constructive comments.

Funding This article was prepared within the framework of the HSE University Basic Research Program in 2019 and funded by the Russian Academic Excellence Project '5-100'.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

[64] However, the correspondence between functional and phenomenal properties may not be necessarily one–one even for non-qualia phenomenal experiences, a topic that has been discussed in identity theory of mind (Smart 2004). For example, the same phenomenal state could be implemented by different physical states (multiple realizability; Bechtel and Mundale 1999).

References

Baars BJ (1988) A cognitive theory of consciousness. Cambridge University Press, Cambridge
Barlow HB (1961) Possible principles underlying the transformation of sensory messages. Sens Commun 1:217–234
Barlow HB (1969) Pattern recognition and the responses of sensory neurons. Ann N Y Acad Sci 156:872–881
Bechtel W, Mundale J (1999) Multiple realizability revisited: linking cognitive and neural states. Philos Sci 66(2):175–207
Berkeley G (1710/1999) A Treatise Concerning the Principles of Human Knowledge. RS Bear
Blakemore SJ, Wolpert DM, Frith CD (2002) Abnormalities in the awareness of action. Trends Cogn Sci 6(6):237–242
Block N (1978) Troubles with functionalism. Minn Stud Philos Sci 9:261–325
Brindley GS, Lewin WS (1968) The sensations produced by electrical stimulation of the visual cortex. J Physiol 196(2):479–493
Chalmers D (1995) Facing up to the problem of consciousness. J Conscious Stud 2(3):200–219
Chalmers D (1996) The conscious mind. Oxford University Press, New York
Chalmers D (2015) Panpsychism and panprotopsychism. *Consciousness in the physical world: Perspectives on Russellian monism*, 246–276
Christoff K, Cosmelli D, Legrand D, Thompson E (2011) Specifying the self for cognitive neuroscience. Trends Cogn Sci 15(3):104–112
Crick F, Koch C (1990) Towards a neurobiological theory of consciousness. Semin Neurosci 2:263–275
Darwin C (1869/2004) On the origin of species. Routledge
David AS, Bedford N, Wiffen B, Gilleen J (2012) Failures of metacognition and lack of insight in neuropsychiatric disorders. Philos Trans Roy Soc B 367(1594):1379–1390
Dehaene S, Sergent C, Changeux JP (2003) A neuronal network model linking subjective reports and objective physiological data during conscious perception. Proc Natl Acad Sci USA 100:8520–8525
Dennett C (1988) Quning Qualia. In: Marcel A, Bisiach E (eds) Consciousness in modern science. Oxford University Press, New York.
Dennett DC (1995) Darwin's dangerous idea. Simon & Schuster, New York
Dennett DC (1999) The Zombic Hunch: extinction of an intuition? Royal Institute of Philosophy Millennial Lecture
Dneprov A (1961) A game. Knowl. Power 5:39–42
Dummett M (1963) Realism. Synthese 52(1):55–112
Fechner GT (1860/1948) Elements of psychophysics. In: Dennis W (ed) Century psychology series. Readings in the history of psychology. Appleton-Century-Crofts, pp 206–213
Fleming SM, Dolan RJ (2012) The neural basis of metacognitive ability. Philos Trans Roy Soc Lond Ser B Biol Sci 367(1594):1338–1349
Fortey R (2011) Life: a natural history of the first four billion years of life on earth. Vintage
Frankish K (2016) Why panpsychism is probably wrong. The Atlantic, 20 September. https://www.theatlantic.com/science/archive/2016/09/panpsychism-is-wrong/500774/
Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RSJ (1995) Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Map 2:189–210
Friston K, Kilner J, Harrison L (2006) A free-energy principle for the brain. J Physiol Paris 100(1–3):70–87

Galton F (1880) Statistics of mental imagery. Mind, os–V 19:301–318
Goff P, Seager W, Allen-Hermanson S (2017) Panpsychism. In: Zalta EN (ed), The stanford encyclopedia of philosophy, Stanford University
Gregory RL (1980) Perceptions as hypotheses. Philos Trans R Soc B 290:181–197
Harman G (1973) Thought. Princeton University Press, Princeton
Hawkins J (2004) On intelligence. Times Books, New York
Helmholtz H (1860/1962) Handbuch der physiologischen optik, vol 3 (Southall J, Trans., ed.). Dover, New York
Husserl E (1913/2012) Ideas: general introduction to pure phenomenology. Routledge
Jackson F (1982) Epiphenomenal qualia. Philos Q 32(127):127–136
Jackson F (2007) The knowledge argument, diaphanousness, representationalism. In: Alter T, Walter S (eds) Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism, pp. 52–64
Kalman RE (1960) A new approach to linear filtering and prediction problems. J Basic Eng 82:35
Kant I (1787/1998) Critique of pure reason. Paul Guyer and Allen W. Wood
Kripke S (1980) Naming and necessity. Harvard University Press, Cambridge
Laertius D (ca. 200/1925) Lives of eminent philosophers. Harvard University Press
Lamme VAF (2006) Towards a true neural stance on consciousness. Trends Cogn Sci 10:494–501
Lau H, Rosenthal D (2011) Empirical support for higher-order theories of conscious awareness. Trends Cogn Sci 15(8):365–373
Levine J (1983) Materialism and qualia: the explanatory gap. Pac Philos Q 64:354–361
Levine J (1999) Conceivability, identity, and the explanatory gap. In: Hameroff SR, Kaszniak AW, Chalmers D (eds) Towards a science of consciousness III: the third tucson discussions and debates. MIT Press, London, pp 3–12
Locke J (1689/1975) Essay concerning human understanding. Oxford University Press, Oxford
MacKay DJC (2003) Information theory, inference, and learning algorithms. Cambridge University Press, Cambridge
Murphey DK, Yoshor D, Beauchamp MSS (2008) Perception matches selectivity in the human anterior color Center. Curr Biol 18(3):216–220
Nagel T (1974) What is it like to be a bat? Philos Rev 83(4):435–456
Nisbett RE, Wilson TD (1977) Telling more than we can know: verbal reports on mental processes. Psychol Rev 84(3):231–259
Popper K (1994/2013) Knowledge and the body-mind problem: in defence of interaction. Routledge
Searle J (1980) Minds, brains and programs. Behav Brain Sci 3(3):417–457
Selimbeyoglu A, Parvizi J (2010) Electrical stimulation of the human brain: perceptual and behavioral phenomena reported in the old and new literature. Front Hum Neurosci 4:46
Shields C (2020) Aristotle. In: Zalta, EN (ed) The stanford encyclopedia of philosophy, Stanford University
Smart JJC (2004) The identity theory of mind. In: Zalta, EN (ed) The stanford encyclopedia of philosophy, Stanford University
Stoljar D (2017) Physicalism. In: Zalta, EN (ed) The stanford encyclopedia of philosophy, Stanford University
Tye M (2006) Absent qualia and the mind-body problem. Philos Rev 115(2):140
Zeman A, Dewar M, Della Sala S (2015) Lives without imagery—congenital aphantasia. Cortex 73:378–380