ChatGPT

Deriving the Existence of Qualia from First Principles

Introduction

Qualia – the raw **subjective qualities** of experience – lie at the heart of the mind-body problem [1]. Philosophers have long debated whether these **phenomenal** aspects of consciousness ("what it is like" to feel pain, see red, taste coffee, etc.) are real, irreducible features of the mind or merely illusions. In this treatise, we construct a systematic argument **from first principles** to demonstrate the necessary existence of qualia. We begin by defining *qualia* precisely and establishing epistemological foundations (e.g. Descartes' *cogito* and self-evident aspects of experience). Building on this foundation, we analyze the concept of consciousness and *subjective experience*, drawing from analytic philosophy and phenomenology. We then progress through classical thought experiments – *Mary's Room*, *philosophical zombies*, *inverted spectra* – which illustrate the irreducibility of qualia. Along the way, we introduce **the Hard Problem of consciousness** and arguments from introspection that highlight an explanatory gap between physical description and subjective feeling [2]. Finally, we address major counterarguments: **eliminative materialism** and **functionalism**, views that attempt to deny, reduce, or explain away qualia. A comparative discussion of dualist, physicalist, idealist, and phenomenological perspectives will clarify how different traditions regard qualia. Step by step, the argument will show that *qualia necessarily exist* as an inherent aspect of conscious mind, even if explaining *how* they arise remains a profound challenge.

Defining Qualia and Conscious Experience

Qualia (singular: *quale*) are commonly defined as the **introspectively accessible, phenomenal aspects of our mental lives** – the *felt* qualities of experience [3]. Whenever you smell a rose, taste a lemon, or feel pain from a pinprick, there is a distinctive *"what it's like"* to undergo that experience. Philosophers often use the term *qualia* to refer to these *subjective* qualities of consciousness. For example, consider looking at differently colored patches: seeing a bright crimson patch feels different from seeing a turquoise patch – each experience has a unique "raw feel" [4]. **Figure 1** illustrates this with three color samples; the *qualitative character* of seeing vermilion, crimson, vs. turquoise is distinct in each case [4]. *Figure 1: Patches of vermilion, crimson, and turquoise – each perceived color produces a different subjective feel or quale.* In general, **qualia constitute the phenomenal character of experience**, i.e. the properties that make *seeing red* feel different from *seeing green*, or that make pain feel different from tickle. These properties are **intrinsic to the experience** and directly accessible through introspection [5]. Crucially, qualia are **private and subjective** – only the experiencer has direct access to how a sensation feels to them. If you and another person both look at a red apple, you each have your own *red* quale; there is "something it is like" for *you* to see red, which might differ (unbeknownst to anyone) from what it is like for them.

In analytic philosophy of mind, qualia are often described in terms of their characteristic features. They are typically held to be **ineffable** (cannot be fully communicated or measured), **intrinsic** (non-relational properties of conscious states), **private** (accessible only to the first-person), and **immediately or incorrigibly apprehensible** (known directly without inference) [6]. Not all definitions insist on all these

features, but this "classic" conception of qualia underlies many debates. Importantly, even a more modest definition – *qualia as just the subjective phenomenal character of experience* – suffices for our purposes. In this broad sense, it is difficult to deny that qualia exist [7]: whenever we attend to our experiences, we find some qualitative *feel* present (be it the sensation of warmth, the bitterness of coffee, or the feeling of anger). Disagreements in philosophy usually concern **which states have qualia and what metaphysical status qualia have**, rather than their existence [1]. By defining qualia as the **felt "what-it's-like" aspect of consciousness**, we have a clear target: to show from first principles that such subjective feelings *must* exist (at least for conscious beings like us), even if their nature and origin are debated.

Foundations: First Principles of Consciousness

Any **first-principles derivation** of qualia's existence must begin with what we cannot rationally doubt – the **data of consciousness itself**. René Descartes famously argued *Cogito, ergo sum* ("I think, therefore I am") as a foundational truth. He observed that even if one doubts everything – the external world, science, even logic – one cannot doubt the fact **that one is doubting (or thinking)**. The very act of thought or experience proves the existence of a thinking subject [8]. This insight provides a **certain foundation**: the reality of one's own mind is indubitable, given that *thought is occurring*. In Descartes' terms, the *cogito* establishes the **existence of consciousness** (a "thinking entity" or mind) as a **first principle** [8].

From the *cogito*, we immediately get the **existence of conscious experience**. To think, doubt, or perceive is to undergo experience; thus the conscious mind is not an empty abstraction but is continually filled with *qualitative* mental events – sensations, thoughts, feelings. Indeed, **conscious experience may be taken as an axiomatic given**: one cannot be mistaken that *one is experiencing something*, at least in general (I might misidentify what I experience, but not the *fact* that an experience is present). This yields our first foundational step: **conscious experiences occur, and their existence is directly known to the subject**. As one philosopher quipped, "what could be more obvious than the existence of conscious experience?" [9]. Any attempt to deny the existence of experience undermines itself, since the denial is itself a conscious act or thought, hence an experience. Thus, we start from the self-evident truth that *our conscious states are real*.

Now, if conscious states are real, we must inquire into their nature. This is where **qualia** enter. Introspection (examining one's own mind) reveals that our conscious states have *particular qualities*. For example, Descartes in his *Meditations* noted that while he could doubt the world, he could not doubt that he *seemed to see* light or *felt* heat – even if those perceptions were illusory, the sensations themselves were real in his mind. We generalize: **each instance of consciousness comes with a distinctive feel**. Whether it's the vivid color of a sunset, the pang of jealousy, or the taste of salt, there is *something that it is like* for the subject to have that experience. This **phenomenal feel is essentially what we mean by "qualia."** Thus, from the **indubitable existence of consciousness**, we derive the **existence of qualitative character**: a conscious mind is not a blank, structureless existence – it is always *somehow* for the subject. Even pure thinking has a "felt" aspect (a train of thought can carry a certain tacit *feel* or imagery). Consciousness, by its very nature, entails *first-person presence*: it *feels like something* to be conscious [10]. We take it as a fundamental principle that **for any conscious being, there is a subjective aspect that accompanies its mental states**. This is essentially an axiom of experience: if I am conscious, qualia (in the broad sense) exist.

To further solidify this foundation, consider the perspective of **phenomenology** (the philosophical study of experience). The phenomenological movement (Husserl, et al.) urges us to "return to the things themselves," meaning to examine experience *as it is given* to consciousness. Bracketing any assumptions about the external world, we find that experience comprises a stream of *phenomena*: sights, sounds, pains,

thoughts, etc., with particular qualities. Phenomenologists treat the **structure of experience** as something to be described precisely, without reducing it away [11]. In doing so, they implicitly affirm that *the reality of these experienced qualities is primary*. We might say phenomenology accepts as a starting point that **experience manifests certain essential features (qualia included), which are to be studied, not denied**. This dovetails with our first-principles approach: the **phenomenal character** of consciousness is a primitive fact – an **immediate given** of the "data of consciousness." Thus, by *epistemic necessity*, **qualia exist**: they are inherent in any conscious state and are known directly to the subject. The task remains to articulate *why* these qualia must be acknowledged and how they resist reduction to other descriptions. For that, we turn to a closer analysis of consciousness and classic arguments.

Consciousness and the "What-It's-Like" of Subjective Experience

What distinguishes a *conscious* mental state from an unconscious one or from mere physical goings-on? The key is often said to be **phenomenal consciousness** – the presence of *qualia*. Thomas Nagel's famous dictum encapsulates this: *an organism is conscious if and only if there is something it is like to be that organism* (to undergo its experiences) [10]. This *something-it-is-like* is precisely the *subjective experience*. For example, there is something it is like to **see red** or to **feel pain**; by contrast, there is nothing it is like to be a piece of rock or (presumably) a thermostatic control system. Nagel's insight brought the **subjective, first-person aspect** of mind to the forefront – a direct reference to qualia. In everyday terms, we might say consciousness is "lit up" from the inside: there is a **first-person perspective in which experiences appear a certain way**. Modern philosophers often label this **phenomenal consciousness**, and they equate it with the presence of qualia (also called experiential or **phenomenal qualities**).

It is useful to distinguish *phenomenal consciousness* from what cognitive scientists call *access consciousness*. **Access consciousness** refers to mental information being globally available for reasoning, speech, and action (the functional, reportable aspects of mind). One could imagine, however, an information-processing system that reports having seen red, reacts appropriately, and so forth, yet *lacks any inner feel*. This hypothetical is precisely a **philosophical zombie**: a being **physically and functionally identical** to a normal human, **but entirely devoid of conscious experience** [12]. Such a being would have all the access mechanisms (it might say "ouch" when poked, avoid injury, etc.) but *no actual feeling of pain* – in other words, all the cognitive processing with **no qualia**. In the zombie, the "lights are off" subjectively, even though behaviorally it is indistinguishable from a person with full consciousness. By contrast, a normal conscious person has both access consciousness *and* **phenomenal consciousness** (the feelings) [12] [13]. This distinction underscores that **qualia are not defined by behavior or function** but by the presence of a subjective point of view.

Philosopher Ned Block uses this distinction to emphasize that current science can explain many *functional* aspects of mind (attention, integration of information, reportability – the so-called "easy problems"), but **the hard part is explaining why and how these functions are accompanied by qualia** [2] [14]. The phrase **"the Hard Problem of consciousness"** (coined by David Chalmers) directly points to the *what-it's-like factor*: Why doesn't all our neural information processing go on "in the dark," without any felt experience? Why is there an inner movie or soundscape of qualia playing out in the mind [15] [2]? This remains deeply puzzling. But for our argument, the critical point is this: **we clearly recognize the presence of the inner qualitative aspect, even if we don't know how it arises**. The reality of *phenomenal consciousness* is evident each time we introspect. As Chalmers notes, even after explaining all brain functions – discrimination, integration, verbal reports – one can still ask: *"Why is all this processing accompanied by an experience?"* [14]. The very asking of this question testifies that we take the existence of the experience (qualia) for granted; it is the

explanandum that needs accounting. In summary, a **conceptual analysis** of consciousness reveals that **subjective experience (qualia)** is an essential feature of mentality. *What it is like* to have a mental state is a property that must be acknowledged alongside any functional or behavioral description. With this in mind, we proceed to examine powerful thought experiments and arguments that have been devised to reinforce the claim that qualia exist and cannot be reduced to other things.

Thought Experiments Demonstrating the Reality of Qualia

Philosophers have devised several compelling **thought experiments** to isolate qualia and show that no purely physical or functional description can capture or eliminate them. These classical scenarios support the claim that **qualia are real and irreducible** aspects of mind. We will discuss three famous examples and what each implies:

- **Mary's Room (The Knowledge Argument):** Frank Jackson's scenario of *Mary* is designed to show that conscious experience involves **non-physical properties or facts**. Mary is a brilliant scientist of vision who knows **everything about color** scientifically – all the physical facts about wavelengths, optics, and neuroscience of color vision – but she has lived her entire life in a **black-and-white room**, never seeing any colors except shades of gray [16] [17]. Mary could, in theory, recite exactly what happens in the brain when seeing red or blue, down to the last neuron. However, when Mary is **released from her monochrome room** and finally sees a red rose for the first time, something remarkable happens. She exclaims, "*So that is what it is like to experience red!*" [18]. Mary learns **something new** – namely, the *experience* of color, the *qualia* of red. This new knowledge is not a new physical fact (she already knew all those), but rather a **direct acquaintance with a quality** she had never felt. The argument thus runs: if Mary had complete physical knowledge yet lacked knowledge of *what it's like* to see red, then **qualia (the redness of red as experienced) are something over and above all physical facts** [16]. In short, **qualia exist** as a kind of fact or property that **escapes a purely physical description**, implying that physicalism is incomplete [19]. Mary's Room suggests that no matter how much you know *about* the mechanisms of an experience, **the actual experience itself (the quale) adds new information** or at least a new mode of knowledge (often called *knowledge by acquaintance*). This vivid thought experiment strongly intuitively supports the existence of qualia: Mary *had* to actually see red to know its feel – demonstrating that the feel is a real feature of the world that wasn't captured by her exhaustive physical knowledge.

- **Philosophical Zombies (Conceivability Argument):** The *zombie* thought experiment, popularized by David Chalmers, asks us to imagine a being indistinguishable from an ordinary human in every physical and behavioral respect (down to identical neural activity) **but lacking any conscious experience**. If such a **"p-zombie"** is even *conceivable* without contradiction, it implies that it is logically possible for all the physical facts about a person to be true while the *phenomenal facts* (the qualia) are absent [12] [13]. This would mean that **qualia are an extra "something"** not guaranteed by the physical configuration – a *"further fact"* of consciousness [13]. The zombie argument is used as a challenge to **physicalism**: if physicalism were true, it should be impossible to have a physical duplicate of you that lacks qualia. But if we find the zombie scenario conceivable (and many do intuitively), then physicalism has not explained or determined qualia. In Chalmers' formulation, **the logical possibility of zombies** shows that no amount of physical description entails the existence of consciousness, hence **qualia exist as a non-derivable aspect** of reality [13]. Another way to put it: when we imagine God creating the world atom by atom (to use a metaphor by Saul Kripke and Chalmers), once all physical facts were set, would any "further act" be needed to bring about

consciousness? If we answer *yes* – God would need to "add qualia" to get conscious beings – then we're admitting qualia are something over and above the physical [20]. The zombie thought experiment dramatizes the intuition that **qualia are real** (we know we have them) yet **one can imagine a world physically like ours *without* them**, meaning they are not redundant concepts. Of course, critics respond that zombies may not actually be possible or conceivable in a coherent way (Dennett, for instance, argues they are *inconceivable* or that we ourselves might as well be zombies if we think that way [21]). Nonetheless, the persistence of the zombie intuition in philosophy of mind (a significant percentage of philosophers consider zombies conceivable [22]) underscores how strongly many thinkers feel that **qualia have a real, undeniable "bite"** that mere physical descriptions lack. The zombie argument, therefore, reinforces the claim: if a being could have all the physical structure of a brain and body and *still* lack the "inner light" of experience, then that inner light (qualia) is a **real feature that is not captured by structural–functional accounts**. It must *exist*, or else there would be nothing obviously missing in the zombie world – yet there *is* something missing: everything that matters about consciousness.

- **Inverted Qualia (Spectrum Inversion):** First discussed by John Locke and later refined by many philosophers, the inverted spectrum scenario asks us to suppose that two people could have **systematically different qualia** even though they behave and function identically. For instance, imagine that what *you* experience as red, *I* experience as green, and vice versa, in a lifelong way – our internal color palette is "inverted." When you look at strawberries (normally "red" objects), you have the sort of experience I normally have when seeing grass; and when I look at strawberries, I have the experience you have when seeing grass [23]. We both call strawberries "red" (having learned the same language and coordinated our behavior), and we both discriminate colors in the same way (we each can tell strawberries and grass apart). Functionally and behaviorally, we're indistinguishable – we make the same color judgments, preferences, etc. Yet by hypothesis, the *qualia* we each have are swapped. This scenario shows that **qualitative experience could vary independently of functional state**. It poses a challenge to any theory that equates mental states purely with physical or functional states (such as behaviorism or certain forms of functionalism). If it's conceivable that *what it's like* for me differs from *what it's like* for you even though we're physically/functionally equivalent in our color processing, then qualia are a distinct aspect that is not captured by the functional description. In philosophical terms, inverted qualia scenarios demonstrate the possibility of **absent or different qualia in functional duplicates**, reinforcing that qualia have a real informational or experiential content above and beyond the functional role [24]. Another version by Ned Block, called *Inverted Earth*, complicates the scenario to avoid some objections, but the core lesson remains: **qualia cannot be purely equated with functional relations or representational content** if such inversions are possible. We each might never discover the difference (since by hypothesis our reports and behaviors are normal to us), but *there would* be** a difference – a difference only discernible from the first-person perspective. This reinforces the reality of qualia: there is a "fact of the matter" about what experience you're having, which is not exhausted by what you say or how you act.

Each of these thought experiments – *Mary, Zombies, Inverted Spectrum* – in its own way drives home the existence of qualia. They show that one cannot explain away or ignore the qualitative aspect of experience without losing crucial information. **Mary** illustrates that no amount of third-person knowledge equals the first-person knowing of a quality [16]. **Zombies** illustrate that one can have all the right physical parts moving and still lack something immeasurably important – the presence of experience [13]. **Spectrum inversion** shows that even the structure of our perceptions (how they relate and cause behavior) might

leave out the actual “colors” of the mind. These are powerful intuition pumps that will inform the more formal arguments about irreducibility and the explanatory gap in the next section.

Introspection, Irreducibility, and the Hard Problem of Consciousness

Having seen intuitive cases for qualia, we now formulate the **core argument**: qualia *exist* and resist reduction because there is an **explanatory gap** between physical descriptions and subjective experience. This is often called **the Hard Problem** – explaining why physical or functional processes give rise to *felt* experience [2]. Here we break down the argument into a more systematic form, including the role of introspection and the alleged irreducibility of qualia.

1. **Direct Introspective Awareness:** When we turn our attention inward, we find certain **immediate qualities present in experience**. For example, if you introspect while looking at a red apple, you are aware of seeing a red patch (the apple’s image). If you then “bracket” the external object and attend only to your *experience* of it, you still encounter a *reddish* visual sensation. Likewise, closing your eyes and focusing on a pain reveals the raw hurt, not just a set of functional reactions. Introspection is our primary epistemic access to qualia. Many philosophers have argued that **introspection provides prima facie evidence** of qualia as intrinsic features of experience – *the redness of the red experience, the painfulness of pain*, etc. It seems obvious that when we feel pain, we *know* the feeling itself. As one philosopher noted, *what it’s like* to undergo a blue visual experience is known to us by simply having the experience [25]. You cannot be wrong that you are currently seeing *some* color or feeling *some* sensation (you might be wrong about whether an external object is truly red or whether your sensation is veridical, but the occurrence of the sensation as such is apparent). **Thus, introspection testifies that qualia are real**: they are **the “fuel” of the phenomenal engine**, so to speak, always present whenever we check. This direct acquaintance is often taken as **self-evident**. Indeed, strong forms of this argument claim qualia are **incorrigible**: if you think you have a certain qualitative experience, you do (at least in that moment) [26]. While this incorrigibility can be debated, the basic point stands: by the very **act of knowing our experiences, we affirm the existence of qualia**. Any theory that says “there really are no qualia” must somehow claim that introspection systematically deceives us – a claim many find hard to swallow, since **conscious experience is arguably the most immediately known reality to us** [9].

2. **Irreducibility and the Explanatory Gap:** Even when we have exhaustive scientific knowledge of a process, a gap remains in explaining its associated qualia. This gap was exemplified by Mary’s case (complete physical knowledge of color vision did not let her deduce the experience of color). Philosopher Joseph Levine coined the term “**explanatory gap**” for the difficulty in bridging the objective and subjective descriptions [27]. In more detail: an objective account might explain **correlations** – e.g. when brain area X is active, the subject reports seeing red – but it doesn’t explain **why** that brain activity *feels red*. From the third-person view, we can trace light of 700 nm hitting the retina, neural impulses through the visual pathway, and activity in color regions of the cortex. We might find a neuron population that specifically responds to red stimuli. But nowhere in that chain is a *redness* that we, as observers, can see. The science describes mechanisms and functions: how signals are processed, how reactions occur. We can even explain the organism’s *behavior*: why it says “I see red,” why it distinguishes red from green in tests. These are the so-called **“easy problems”** (though they are technically challenging, they are considered *easy* in principle because they are

about functionality and mechanism [28] [2]). The **Hard Problem** remains: *why does all that processing generate an experience of redness?* Why isn't the processing "dark"? This is not a scientific ignorance that more data could simply fill in – it's a conceptual gap. There seems to be **no logical derivation** from knowing everything about neural circuits to knowing what a subjective experience will be like [2]. Philosopher David Chalmers expresses it thus: even after we account for every cognitive function (discrimination, integration, report, etc.), we can still sensibly ask, *"Why are these functions accompanied by experience – why* aren't *we zombies?"* [14]. The continued meaningfulness of that question indicates that **qualia are something extra that our theories leave unexplained** [2]. In philosophical terms, *qualia do not logically supervene on the physical* – fixing all the physical facts doesn't obviously fix the facts about qualia. This irreducibility claim doesn't *prove* dualism or anything, but it shows that **qualia have a special status**: they resist the standard forms of reduction that work for other phenomena (like life or heat, which *did* turn out reducible to molecular mechanisms). There is a **huge intuitive difference** between explaining how a brain discriminates colors and explaining why *red looks the way it does to someone*. The latter is an *additional fact* in need of explanation [13]. Therefore, by *abduction* (inference to the best explanation), one concludes that **qualia are real features** that are not captured by the physical story, rather than illusions. They are *the thing that needs explaining* in the Hard Problem, which is itself evidence of their reality: if there were no qualia, there would be no Hard Problem – nothing "mysterious" to explain, just functions. But there *is* something mysterious, precisely because we know qualia so intimately yet cannot see how to integrate them with physical theory [2].

3. **The Inescapability of the First-Person Perspective:** Another irreducibility argument comes from the **first-person perspective** being irreducible to any third-person description. Thomas Nagel's *bat* example is apropos: we can know every objective fact about bat sonar, yet *there is something we cannot know*: what it is like *for the bat* to experience the world [2]. This "for-the-subject" aspect is unique. An objective physical theory can describe brains and signals but will always do so in third-person terms (occupying such-and-such positions, obeying equations, etc.). The **subjective feel** is simply not present in those terms – it's only present from the *inside*. This suggests that any objective ontology is, at best, incomplete when it comes to mind. **Qualia exist at the first-person level**, and capturing them may require a different approach or expansion of our ontology. Indeed, some philosophers (e.g. proponents of **panpsychism** or **dual-aspect theories**) have argued that the intrinsic nature of physical entities might include experiential aspects to close this gap [29]. We will not delve into those solutions here; the point is that the **reality of the first-person viewpoint** is unavoidable and brings with it the reality of qualia. The fact that we can imagine zombies or that we feel an explanatory gap is itself a result of comparing the first-person knowledge of qualia to the third-person knowledge of mechanism. Thus, irreducibility underscores existence: since qualia *cannot be derived away*, they must be acknowledged as part of what fundamentally is.

4. **Introspection vs. "Transparency":** A brief caveat: Some philosophers (often **representationalists**) have argued that introspection doesn't actually reveal intrinsic qualia – instead, we supposedly "look through" our experiences to the world (this is the **transparency** of experience) [30] [31]. For example, when you introspect on your visual experience of a red apple, you *seem* to only notice the apple's redness, not a mental redness; thus, these philosophers claim that the only qualities we encounter are properties of external objects as represented, not properties of experience itself [30] [31]. If true, this would imply that qualia (as intrinsic mental paint) might be an illusion; what we have are *representational contents*. However, even this view concedes that there is *something it's like* for you to see the apple – it's just attributing that "something" to world properties represented in your mind

rather than ghostly mental paint. The **qualitative character still exists**; the debate is whether it is a property of a mental state or of an intentional content. Moreover, opponents of the transparency thesis argue that on careful reflection we *can* notice the subjective aspect (for instance, the difference between veridical vision and hallucination isn't in the external world – it's in our experience). Regardless, the transparency debate doesn't eliminate qualia; it merely relocates or redescribes them. The **obvious fact remains**: *experiences differ in how they feel*, and this feeling is directly known to us (whether we call it an intrinsic property or a representational content). Thus, the **existence** of qualia is not denied by transparency advocates; instead, they try to reconcile qualia with a physicalist view by saying qualia are just the way external properties are represented to us [31]. But if anything, this reiteration shows how persistent the notion of qualia is – even theories that want to "subtract" spooky qualia end up having to account for the qualitative, *felt* differences in experiences, albeit by another name.

In sum, by appealing to **introspection**, **logical arguments (knowledge gaps, conceivability)**, and the **Hard Problem**, we have built a case that **qualia are real, fundamental aspects of consciousness**. We started from the indubitable presence of experience, noted the distinctive qualitative nature present, and saw that no reductive explanation currently removes the need to accept those qualities as real. The *existence* of qualia stands firm; what remains open is the nature of their existence (are they non-physical properties? Or identical to complex physical states in a way we don't yet understand? etc.). Before concluding our derivation, we will contrast a few major philosophical perspectives on qualia to show how they differently accommodate or deny these conclusions, and address head-on the **views that attempt to deny or "dissolve" qualia**.

Philosophical Perspectives: Dualism, Physicalism, Idealism, Phenomenology

Qualia have been a flashpoint in debates between various **philosophical traditions**. To strengthen our argument, it's helpful to see how *different worldviews* conceive of qualia – and why denying qualia's existence is problematic. Below we compare perspectives:

- **Cartesian Dualism (Mind-Body Dualism):** In the Cartesian tradition, mind and matter are fundamentally different substances (res cogitans vs. res extensa). Qualia, as features of conscious experience, fall squarely on the *mind* side of this divide. For René Descartes and many subsequent dualists, the **existence of the mind's qualitative states** is undeniable – in fact, Descartes used *pain* and other sensations as examples of things known *more certainly* than anything physical (you can doubt the presence of fire, but not the feeling of heat you seem to feel) [8]. In dualism, **qualia are real properties of mental substance**. They are typically seen as *non-physical* (since the mental substance itself is non-material). This view faces the classic interaction problem, but it *at least* cleanly accepts qualia: the feel of pain is an irreducible mental property. We might call this a **qualia-friendly** view – qualia are expected, even central, since the mind's whole essence is to think and feel. Many dualists cite precisely the sort of first-person evidence we have outlined as support: **physical mechanisms alone can't account for qualia, so the mind must have a non-physical aspect or substance in which qualia inhere** [32] [33]. Our derivation from first principles resonates with dualism's starting point (the certainty of consciousness). However, dualism doesn't *explain* qualia so much as place them in a separate ontological category. Nonetheless, for a dualist, the **necessary existence of qualia** is a given – they are part of the basic furniture of reality (mental properties).

- **Physicalism (Materialism):** Physicalism is the view that ultimately only physical things exist, and mental states are either identical to physical states (in reductive physicalism) or at least wholly determined by them (non-reductive physicalism). This tradition faces what we have described as the *Hard Problem*. Early identity theorists (like J.J.C. Smart) tried to identify sensations with brain processes (e.g. "pain is C-fiber firing"). While this may be true in some sense, it doesn't immediately illuminate why C-fiber firing *feels like pain*. More sophisticated physicalist approaches, such as **functionalism**, understand mental states in terms of their causal roles – what they do rather than what they feel like. Functionalists might say a pain is whatever state plays the "pain role" (caused by injury, produces avoidance behavior, etc.). Classic functionalism was initially ambivalent or silent about qualia, treating the *internal feel* as irrelevant or as automatically accounted for if the functional role is present. But as we saw, thought experiments like inverted spectrum pose serious challenges: two systems could be functionally identical yet experience different qualia, implying qualia are not fixed by functional role [24]. This put pressure on functionalist theories: either deny the possibility of such inversions or extend the theory to account for qualia. Some physicalists (especially **eliminative materialists** and **illusionists**, discussed below) go so far as to claim that qualia *do not exist* or are a kind of cognitive illusion. Other physicalists take a reductionist but qualia-positive route: they aim to eventually *explain* qualia in terms of brain activity patterns (for instance, suggesting that a specific neural oscillation or information integration corresponds to the feeling of red). Contemporary neuroscience seeks the **neural correlates of consciousness (NCCs)** – minimal brain states necessary for specific experiences [34]. While this research is valuable, it hasn't yet explained *why* those NCCs give rise to qualia. Physicalists often have to bite the bullet on the Hard Problem, maintaining that *in principle* science will close the gap. Some propose that we need new concepts or even new fundamental properties (as in **Russellian monism** or **panpsychism**) to account for qualia without abandoning physical monism [29]. In summary, physicalist views **vary**: moderate physicalists accept qualia exist but strive to reduce or explain them physically; extreme ones deny qualia as normally conceived. The very need for such extremes or explanatory contortions in physicalism, however, can be seen as evidence from our first-principles view that **qualia are indeed "something extra" that this framework struggles to accommodate**.

- **Idealism:** At the opposite end of the spectrum from physicalism is **idealism**, the view that reality is fundamentally mental. For an idealist (e.g. George Berkeley), matter as normally conceived doesn't exist; rather, what we call physical objects are collections of perceptions or ideas in minds. In Berkeley's famous slogan, *esse est percipi* – "to be is to be perceived" [35]. Under subjective idealism, **everything we ever "know" of the world are qualia** – colors, sounds, sensations that exist in the mind [36]. If you remove all percipient minds, you remove the world of appearances; there is no mind-independent, qualia-free "substance" out there, according to this view. Thus, **qualia have the strongest reality in idealism**: they are not epiphenomena or byproducts, they are the very stuff of the world. The external world is essentially a shared dream of qualia (possibly coordinated by a divine mind, in Berkeley's theory). Our argument from first principles aligns with idealism to the extent that it prioritizes the data of experience. Indeed, Berkeley started from reflections not unlike our own: you never encounter "matter" directly, only your perceptions (qualitative experiences) of it [37] [38]. He concluded that as far as we can meaningfully speak, **"to exist is to be perceived or to perceive."** From this standpoint, *denying qualia* is nonsensical – there is nothing else *to* reality except qualia (and minds). While most modern philosophers don't embrace full-blown idealism, this view usefully highlights a point: *whatever else we doubt, we cannot doubt experiences*. An idealist simply takes that insight to its limit, making qualia primary and matter secondary or nonexistent. Our derivation doesn't require endorsing idealism, but it certainly draws strength from the idea that

qualia are the primitives of our epistemology – the first things given – which even an idealist and a dualist and many physicalists could agree on.

- **Phenomenology and Existential-Phenomenology:** The **phenomenological tradition** (Husserl, Heidegger, Merleau-Ponty, Sartre, etc.) does not usually talk of "qualia" per se, but it is entirely focused on **the structure of experience**. Husserl introduced methods like *epoché* (bracketing the natural attitude) to analyze experience without bias. In doing so, he effectively treats **phenomena (appearances in consciousness)** as the starting point – much as we have. For phenomenologists, **conscious experience has an inherent structure and content** (the *phenomenology* of it) that can be described without reduction. They explore aspects like intentionality (the aboutness of experience) and the lived body, but they certainly acknowledge *sensory qualities, feelings,* etc. as part of the descriptive tapestry. In some ways, phenomenology offers a **systematic first-person science of qualia** (though they wouldn't use that term) – by rigorously describing how things *appear* to us, they implicitly catalog qualia. One example: phenomenologists distinguish between different modes of experience – perception, imagination, memory – noting the qualitative differences in how an imagined red vs. a perceived red feels, etc. They would resist any view that these lived-through qualities are unreal. In fact, one might say phenomenology is about **taking qualia seriously** and understanding them in context (how they structure our experience of the world). This approach complements the analytic arguments by ensuring we don't lose sight of the **richness and reality of the subjective "lifeworld."** It also addresses a potential worry: one could argue our focus on discrete qualia (red, pain, etc.) is too atomistic, whereas in lived experience qualia come as part of a unified, meaningful experience. Phenomenologists would agree that experiences are holistic – *qualia never float free; they are always part of the experience of a world*. But far from eliminating qualia, this simply situates them. The warmth qualia of holding a cup of coffee, the color qualia of sunrise, the mood qualia of a melancholic afternoon – all these are **parts of the fabric of experience** that phenomenology seeks to describe. Thus, phenomenology provides further validation: it starts with the conviction that **experience (with all its qualities) is real and the primary thing we can analyze**. Any attempt to deny that would, in this view, amount to a failure to perform the phenomenological *reduction* properly (i.e. sneaking in an unfounded theoretical bias that "experience isn't real").

- **Property Dualism and Panpsychism (Middle-ground views):** There are also hybrid views attempting to reconcile qualia with a scientific worldview without going full Cartesian or eliminativist. **Property dualism** holds that there is only one kind of substance (usually physical), but it has two distinct kinds of properties: physical properties and mental (qualitative) properties. On this view, **qualia are real properties** that *supervene* on physical structures but are not reducible to physical properties. They might be seen as emergent. Another family of views, like **Russellian monism** or **panpsychism**, posits that at the fundamental level, matter has an "inner" aspect that could be proto-mental or experiential in nature [29] [39]. In Russellian monism (inspired by Bertrand Russell), physics describes only the *relational, structural* features of matter, not its true intrinsic nature; that intrinsic nature might well be consciousness or proto-consciousness. Thus, human consciousness with full-fledged qualia could arise from combinations of tiny "micro-experiences" in particles, or from underlying qualia-like properties of fundamental entities [40] [41]. These views are motivated by exactly the considerations we've discussed: **the desire to acknowledge qualia as real and fundamental** (in some sense) while still maintaining a monistic ontology (not two separate substances). They effectively grant that classical physicalism as "matter without mind" is incomplete: something must be added or recognized to give matter the capacity to produce mind. Our earlier

inference – that physical descriptions leave something out – is the starting point for these theories. By positing that qualia (or proto-qualia) are **ubiquitous and basic**, they ensure that *when* matter organizes a brain, the qualia are there, because they were always part of matter's story. Whether or not one finds these plausible, it's telling that serious philosophers and even scientists consider such moves necessary to deal with qualia. It underscores that **qualia's existence is taken as non-negotiable; what's debated is how to fit them into our worldview**. The very proliferation of theories (dualism, panpsychism, etc.) to account for qualia is testament to the fact that qualia *are* recognized as needing accounting for – which implicitly means their existence is a datum any adequate theory must handle.

In surveying these traditions, a pattern emerges: **the existence of qualia is far less controversial than the explanation of qualia.** Dualists, idealists, phenomenologists, property-dualists, panpsychists – all firmly assert qualia (they just disagree on ontology). Even many physicalists, when pressed, admit that the *phenomena* of consciousness are real – they just think eventually those will be explained physically, or else they recast them in functional terms. The only truly qualia-denying views are the eliminativist or extreme illusionist ones, to which we now turn. Seeing how one might attempt to deny qualia will further solidify why those denials are not compelling in light of our first-principles reasoning.

Counterarguments and Qualia-Denying Views (Eliminativism and Illusionism)

Despite the strong intuitive and philosophical case for qualia, some theorists have argued that qualia *don't really exist* – or at least, that our ordinary conception of qualia is so mistaken that we should jettison it. These views arise largely from a commitment to a certain form of scientific materialism which finds qualia intractable. If something doesn't fit the preferred ontology, one radical solution is to declare it an "illusion." **Eliminative materialism** in philosophy of mind holds that our common-sense mental concepts (like "belief," "pain," or "qualia") might be as misguided as obsolete scientific concepts like "phlogiston" or "the ether." In an eliminativist view, it's conceivable that a mature neuroscience would not include anything like "qualia" in its theory, and thus we'd conclude qualia were a mistaken construct of folk psychology [42] [43].

One version of this is to deny *consciousness* (in the phenomenal sense) outright – saying that there is no inner experience, only information processing. Philosopher Georges Rey and others have advocated such **consciousness eliminativism**, arguing that the very notion of consciousness might be ill-defined and unnecessary [44] [45]. As incredible as it sounds ("how can one deny that conscious experience exists?" [46]), they suggest that what we *call* conscious experience might be a kind of conceptual confusion or a byproduct of language. The motivation is often: if science can't objectively detect or define qualia, maybe qualia aren't objectively real. Another influential thinker, Daniel Dennett, has famously argued in *Quining Qualia* (1988) and *Consciousness Explained* (1991) that the traditional notion of qualia (as private, ineffable mental "feels") is muddled and should be discarded [6]. Dennett doesn't deny that we *think* we have inner experiences, but he contends that when you examine what people actually do and say, there's no coherent remainder that is an irreducible quale – instead, all that exists are **functional brain states and their dispositions** (what he calls the "multiple drafts" or content discriminations in the brain). In effect, Dennett's view is that **qualia are a sort of user-illusion** created by our brain's informational processes; the brain makes us "privy" to certain contents and we report them, but there is no extra ineffable private property beyond those reports.

A more recent movement along these lines is **illusionism about consciousness**, advocated by philosophers like Keith Frankish. Illusionists maintain that consciousness (specifically, *phenomenal consciousness*) is not what it seems – our brains trick us into thinking there is a private theatre of qualia, whereas in reality there are just cognitive representations, functions, and reactions [47]. To an illusionist, when we introspect and feel absolutely sure we have qualia, we are simply under the spell of an evolutionarily crafted introspective *illusion*. Frankish boldly says we are *wrong* in thinking we undergo experiences with qualitative properties [48] – we are information-processing machines that have gotten into the habit of telling a story about a "mind" with qualia. This strong form of illusionism amounts to a form of eliminativism: deny qualia *tout court* (at least qualia in any sense beyond information encoding). Proponents acknowledge it *sounds* absurd ("what could be more obvious than the existence of conscious experience?" as even illusionists admit [9]), but they argue that introspection is not a reliable guide – it **misleads us** into positing inner qualities when in fact all that exists are physical/functional states [9].

How might one justify such a counterintuitive claim? Typically by pointing out that many intuitive ideas (like the motion of the sun around Earth) turned out to be illusions produced by perspective. They suggest consciousness might be another case where nature "fools" us. Perhaps there is a functional explanation of why we *report* qualia even though none actually accompany the processing – e.g. a theory of introspection that shows it is possible to have brain circuits that assert they have ineffable qualia even when they don't [49]. Illusionists like Frankish propose to develop such models. The burden is heavy: they must explain how and why the brain would produce *such an elaborate deception*, and why this is more plausible than just saying qualia exist. As it stands, **illusionism is a minority (but vocal) position**, often criticized as "denying the datum" or "throwing the baby out with the bathwater" [50] [51]. Indeed, many find eliminativism about consciousness to be *too strong a reaction*, one that **seems to fly in the face of our basic epistemic certainty** [52]. Recall Descartes: even if an evil demon deceived you about the external world, it could not deceive you into merely *thinking* you have experiences when you actually have none – for the deception itself would be an experience. The eliminativist has to say, in effect, that we could be "phenomenal zombies" that merely think they aren't zombies. This is a logically possible scenario, but as some critics point out, it verges on self-defeating: if *all* our introspective convictions are that wrong, how can we trust any cognition at all, including the cognition that leads the eliminativist to that theory?

Our first-principles approach gives a clear reply to eliminativism/illusionism: **the existence of qualia is more certain than the theories that deny them**. We directly know pain hurts; to then claim "actually, nothing hurts, you just have a disposition to avoid certain stimuli and claim you feel pain" seems to conflate the public description with the private reality. It also **begs the question** – it assumes from the start that only the third-person, functional story is real, which is exactly what is at issue. Moreover, eliminating qualia does not really solve the Hard Problem – it just declares it a non-problem by fiat ("no consciousness, no problem" [53]). This "solution" might remove an embarrassment for materialism, but at the cost of denying the very subject matter that materialism was supposed to explain. As philosopher Galen Strawson quipped, the crazy thing is not that we *have* consciousness (qualia) – the crazy thing would be denying we have it, since nothing is more evident in our lives. Thus, **most philosophers reject eliminativism about qualia as implausible**, holding that it is "highly counterintuitive to deny that consciousness exists" and that we should find a view that *accounts* for consciousness rather than eliminating it [50]. Even those sympathetic to a reductionist approach concede that we must somehow keep consciousness in the picture, perhaps by reconceiving it, but not denying it entirely [54].

Functionalist Counterarguments: A related, though less extreme, counterargument comes from functionalist or reductionist theories which attempt to *explain* qualia in wholly functional terms, thereby (if

successful) dissolving any mystery. For instance, the **Global Workspace Theory** (Baars, Dehaene, etc.) suggests that a mental state is conscious if it is broadcast to a broad network enabling global access [55]. If someone asks, "why is that state conscious (has qualia)?," the theory's answer is, "because it's in the global workspace – that's what being conscious *means*" [56]. In other words, these theorists build **the definition of consciousness into a functional architecture**. Similarly, **first-order representationalists** (like Dretske or Tye) claim that a conscious experience is just a representation of the world that is poised to inform the organism, and its *qualia* are simply the contents represented (the redness, loudness, etc., as properties of external or bodily objects) [57] [58]. On such views, if we ask "why do we have qualia?", the answer might be: "We have *information* about color, sound, etc., and when that information is accessible in certain ways, that *is* having a qualitative experience." These accounts strive to **show that qualia are not extra mysterious additions, but just another way of describing certain functional states**. If fully successful, they would "dissolve" the intuition of something non-physical by making it clear how *what-it's-like* is just *what-it-does*.

However, many critics argue that these accounts either change the subject or leave something out. When a functionalist theory says "that's what being conscious *means*," one can retort: it only *relabels* the problem. We can still ask: *why does a globally broadcast representation feel like something to the organism that has it?* The theory might answer in a circular way ("because that availability constitutes feeling"), which not everyone finds satisfying. It often seems that such theories only target the "easy problems" of correlation and function [28] [2], and *stipulate* that the hard part is solved by definition. Our argument so far indicates that **something would still be left unexplained** in a purely structural account – namely, the intrinsic *appearing* of those representations to a subject. If a theory can truly account for that, then it would be a breakthrough; but to date, none is widely accepted as doing so. In fact, some defenders of functionalism acknowledge that to secure their view, they must deny that qualia as commonly conceived (intrinsic, possibly non-functional properties) exist – essentially sliding into the illusionist stance [59]. For example, if you define consciousness as "what happens when the brain integrates information in way X," you are basically saying qualia = integration X. If someone then says "I can imagine integration X without the feel," you might be forced to say "well, that imagined 'feel-less' integration isn't actually possible, or you're conceptually mistaken." Thus functionalists often end up arguing that **inverted spectra or zombies aren't really conceivable** given the physical/functional facts – a fair move, but one that begs the question against the conceivability intuitions rather than resolving them.

In short, **functionalism's counterargument** is to insist that *qualia = certain functional patterns*, and that our intuitions to the contrary are either wrong or not well-founded. This is a less extreme stance than saying "qualia don't exist at all," but it still downplays the unique nature of qualia by trying to identify them with something else. From a first-principles perspective, we should be cautious: such identification only works if it truly captures all aspects of the phenomena. As of now, it remains controversial whether any functional account has captured *why a given pattern feels the way it does*. Our step-by-step derivation has highlighted precisely that *uncaptured aspect* as the essence of qualia's existence. Therefore, while functionalist theories are important in showing how consciousness might be integrated with brain function, they do not actually refute the existence of qualia – they either assume it under a different guise or risk explaining it away in a manner many find incomplete.

Summary of Counterarguments: Eliminativism and strict functionalism attempt, in different ways, to defuse the problem of qualia by denying its premise (that qualia are irreducible). Eliminativists straightforwardly deny the existence of qualia (or redefine "consciousness" such that qualia are not included) [44] [45]. Functionalists don't deny that experiences occur, but they deny there is any extra *feely* property beyond functional organization – effectively claiming qualia are just how certain processes *inform*

the system. In evaluating these, it's worth noting our initial first principle: the **reality of experience**. Eliminativism, by contradicting this, faces a steep uphill battle: it asks us to trust abstract theory over phenomenological reality. Most find this unreasonable: *pain hurts, no matter what theory says*. Functionalism, meanwhile, must be careful not to slide into an **analytic limbo** where the definition of consciousness solves the problem by excluding the troublesome part. The impasse these counterarguments often reach tends to reinforce the idea that **qualia's existence is more secure than any theory that tries to eliminate them**. As the Stanford Encyclopedia of Philosophy notes, such radical positions "seem much too strong," given how basic qualia are to our conception of mind [50]. The far more common approach (even among physicalists) is to accept qualia as real and then seek a theoretical framework to accommodate them, rather than deny their existence.

Conclusion

Beginning from undeniable **first principles** – *the certainty of our own conscious experience* – we have systematically derived the conclusion that **qualia exist** and cannot be dismissed. We defined qualia as the **phenomenal, subjective aspects** of mind, the "raw feels" known through introspection [3]. We grounded our argument in the self-evidence of consciousness: as Descartes showed, one's own thinking and perceiving cannot be doubted, and hence the *mind's existence with its conscious states* is fundamentally certain [8]. Within those conscious states, we directly apprehend qualitative features (the **what-it's-like** character), which form the very substance of our mental life. From this basis, we examined conceptual analyses and thought experiments demonstrating that these qualitative features (*qualia*) are **indispensable** and **irreducible**. Whether through Mary gaining new knowledge of color [16], or the logical possibility of a zombie lacking inner light [13], or the notion of inverted qualia escaping functional capture [24], we saw that **no description of the world is complete without accounting for the first-person feel of experience**. David Chalmers' Hard Problem and Levine's explanatory gap reinforced that intuition in theoretical terms: even a perfect scientific understanding of function leaves a **gap** where *subjectivity* lies [2].

Our journey through various philosophical traditions showed that **qualia are acknowledged in all serious approaches to consciousness**, albeit interpreted differently. *Dualists* treat them as fundamental mental properties; *physicalists* struggle to explain them but largely concede they are the phenomena to be explained; *idealists* elevate them to the primary reality of the world; *phenomenologists* study them as the essence of how the world appears to us. The only voices denying qualia (eliminative materialists and extreme illusionists) demand we reject the most immediate evidence of our existence – a move that, upon scrutiny, seems less parsimonious or plausible than accepting that **our experiences are indeed real** [52]. After all, a theory that says "nobody is actually conscious, we just think we are" has arguably *more* explaining to do (to account for that grand delusion) than one that simply takes consciousness as a given reality and tries to integrate it. Our first-principles approach vindicates this common-sense stance: the *existence* of qualia is not what's in question – it's how they fit into the natural order. In short, **qualia necessarily exist wherever consciousness exists**; they are, by definition, the *manifestations* of having a conscious perspective.

Far from being an airy philosophical posit, qualia are the **only "pieces of reality" we cannot doubt** – the redness I see, the pain I feel, these are **experiential axioms**. Any further philosophical or scientific inquiry must accommodate them. A physicalist may attempt to identify them with brain states, a dualist may place them in a separate substance, a panpsychist may distribute them across fundamental particles – but none of these moves negate the central point: *qualia are there to be dealt with*. They are, in a sense, the **starting**

point of any theory of mind (even if some theories attempt to clutch them only indirectly). Our systematic derivation has, step by step, established this on firm ground.

In closing, we acknowledge that **explaining qualia** (the *how* and *why* behind them) remains an open problem. But recognizing their **unavoidable existence** is a crucial and non-negotiable step. As we've shown, attempts to wish them away run into logical or phenomenological absurdities. Thus, we conclude that **qualia exist as an inherent and necessary aspect of conscious beings**, an existence confirmed by immediate self-awareness and resilient against all forms of philosophical doubt. In the grand scheme, this insight reaffirms the profound mystery and importance of consciousness: there is something undeniably real about the *first-person viewpoint*. Any "Theory of Everything" that a future science might formulate will be incomplete if it leaves out that inner universe of qualia [2]. Our derivation from first principles ensures that *that* inner universe is never ignored, for it is, quite literally, **the only universe we directly know**.

Sources: The arguments and examples above draw on a broad philosophical literature. Key references include the Stanford Encyclopedia of Philosophy entries on *Qualia* [3] [6], *Mary's Knowledge Argument* [16], and *Zombies* [12] [13], the Internet Encyclopedia of Philosophy on the *Hard Problem of Consciousness* [2] [14], as well as classic papers by Thomas Nagel on *what it's like* [10], Frank Jackson on *epiphenomenal qualia*, David Chalmers on *the hard problem*, and Daniel Dennett on *quining qualia*. These sources support the claims that (1) qualia are the subjective, phenomenal features of experience, (2) their existence is evident and widely acknowledged, and (3) they pose a serious challenge to reductive explanation [1] [2]. The comparative views and counterarguments were also informed by these sources, illustrating the spectrum from qualia-affirming to qualia-denying positions [44] [47]. Each citation in the text corresponds to a specific supporting passage in those works. Together, they provide a rich intellectual backdrop for the derivation presented here, which remains faithful to the evidence of *experience itself* as the ultimate proof of qualia.

---

1 3 5 6 7 9 17 18 24 25 26 29 30 31 39 40 41 47 48 49 Qualia (Stanford Encyclopedia of Philosophy)
https://plato.stanford.edu/entries/qualia/

2 14 15 27 28 42 43 44 45 46 50 51 52 53 54 55 56 57 58 59 Hard Problem of Consciousness | Internet Encyclopedia of Philosophy
https://iep.utm.edu/hard-problem-of-conciousness/

4 23 Inverted Qualia (Stanford Encyclopedia of Philosophy)
https://plato.stanford.edu/entries/qualia-inverted/

8 Cogito, ergo sum - Wikipedia
https://en.wikipedia.org/wiki/Cogito,_ergo_sum

10 11 Phenomenology | Internet Encyclopedia of Philosophy
https://iep.utm.edu/phenom/

12 13 21 22 Philosophical zombie - Wikipedia
https://en.wikipedia.org/wiki/Philosophical_zombie

16 19 Qualia: The Knowledge Argument (Stanford Encyclopedia of Philosophy)
https://plato.stanford.edu/entries/qualia-knowledge/

20 32 33 Zombies (Stanford Encyclopedia of Philosophy)
https://plato.stanford.edu/entries/zombies/

34 Qualia - ScienceDirect.com
https://www.sciencedirect.com/science/article/pii/S096098221200320X

35 36 37 38 George Berkeley’s Subjective Idealism: The World Is In Our Minds | Philosophy Break
https://philosophybreak.com/articles/george-berkeley-subjective-idealism-the-world-is-in-our-minds/