Consciousness thought experiments with Non-Referential Terms

Abstract

This note (it is not a full-fledged academic paper) introduces a novel approach to classic thought experiments in consciousness studies through the incorporation of non-referential terms—symbols that present experiences directly rather than referring to them. By analyzing the Hard Problem, Knowledge Argument, Philosophical Zombies, and Spectrum Inversion thought experiments using both referential terms (like "blackness") and non-referential terms (like █), the paper reveals that many apparent philosophical puzzles arise from conflating referential descriptions with direct presentational experiences. The analysis shows that attempting to formulate these thought experiments using non-referential terms often requires instantiating the very experiences in question, creating self-referential paradoxes that solve or dissolve the original problems.

Introduction

The history of philosophical thought experiments in consciousness studies has been dominated by referential terms - words and symbols that point to or represent something beyond themselves. From the earliest writing to contemporary philosophers, from Eastern wisdom traditions to modern analytic thought, our conceptual frameworks have operated within a universe of reference where terms invariably stand for or indicate something else. This paper introduces a radical departure from this tradition: the formal incorporation of non-referential terms into our analysis of classic thought experiments in consciousness studies.

Consider the term "blackness." When we use this word, it refers to a particular quale - the subjective experience of the color black. One might be tempted to interpret █ as simply another referential term for "blackness," another way of indicating or pointing to the same experience. However, this interpretation would fundamentally misunderstand the nature of non-referential terms in the calculus of qualia. Following Merriam and Habeeb (2024), we can express this crucial distinction through the inequality:

blackness ≠ █

This inequality holds because these terms function in fundamentally different ways. While "blackness" refers to the experience, █ presents the experience directly. It does not point to or represent anything - not even itself. Its presentation is its entirety. This is not merely a notational convenience or alternative symbolism - it represents a fundamental expansion of our conceptual framework to include terms that present rather than refer.

The distinction between referential and non-referential terms is not merely definitional but represents a paradigm shift in how we conceptualize consciousness and its relationship to reality. When applied to classic thought experiments in consciousness studies, it reveals new dimensions of these problems and suggests novel solutions that were previously invisible within purely referential frameworks.

This note looks at the consequences of non-referential terms for the Hard Problem, Knowledge Argument, Philosophical Zombie, and Spectrum Inversion thought experiments.

The Hard Problem of Consciousness

Chalmers's (1996) hard problem asks why physical processes give rise to conscious experience at all. The traditional formulation suggests an unbridgeable explanatory gap between physical descriptions and subjective experience. This has been central to debates about the mind-body problem and the nature of consciousness.

Non-referential terms transform this problem in fundamental ways. The gap appears unbridgeable when we try to derive referential "consciousness" from physical descriptions. But perhaps █ isn't derived at all - its presence might be fundamental rather than emergent. The first statement of the Hard Problem that is relevant to the discussion:

(1) There is an explanatory gap between body and mind (qualia).

We analyze this as using two referential terms, “body” and “mind (qualia)”, and seems to have no possible answer. We can state a more precise Hard Problem like this:

(2) There is an explanatory gap between body and blackness.

And finally

(3) There is an explanatory gap between blackness and █.

The former term is referential and the latter term is non-referential. It is presentational.

This formulation is best as it presents the endpoint of one half of the answer for which an explanation is needed, namely, any answer to the Hard Problem must include █ itself. In (1) and (2) we have talked *about* a Hard Problem but didn’t actually give a Hard Problem. Whether (1) and (2) have an answer or not does not concern us. The question we want answered is (3).

The formulation of the answer crossing the gap must include █ and blackness (not “blackness”). We understood the blackness part of the previous sentence by having a specific subjective experience. Thus, what we want to do is bridge the gap between the subjective experience of having that understanding and subjective experience █. The question becomes: is there a path of experiences that connect them, and is it implementable? If so, you have an answer. If not, not.

The Knowledge Argument

Mary the color scientist knows all the physical facts about color vision but lives in a black and white room. What goes on as she steps out into a colorful world? This paper’s output is only in black and white, so we’ll state the Knowledge Argument as:

(4) Mary knows all of the physical facts about color, including the color black, but is raised in a grey-and-white room. In stepping out into a world that contains blackness, does she learn something new?

We replace this with

(5) Mary knows all of the (referential) physical facts about color, including the color black, but is raised in a grey-and-white room. In stepping out into a world that contains blackness, does she learn something new?

Which gets replaced with

(6) Mary knows all of the (referential) physical facts about color, including the color black, but is raised in a grey-and-white room. In stepping out into a world that contains █, does she learn something new?

This latter is a Knowledge Argument (question). Neither (4) nor (5) are. So it is (6) we want an answer to.

This formulation makes it clear that what Mary lacks is not referential knowledge about blackness, but rather the direct presentational experience of █ itself. In her grey-and-white room, Mary possessed complete knowledge of all physical and referential facts about "blackness," yet she had never encountered █. Upon stepping out, she experiences █ for the first time - not as new referential knowledge about "blackness," but as her initial encounter with its actual presentational meaning.

The profound implication of this distinction lies in the nature of Mary's knowledge. Her comprehensive understanding within the room was entirely referential. Since █ cannot be reduced to or derived from referential knowledge alone, as demonstrated in this paper, Mary gains something genuinely novel upon leaving - not additional facts about blackness, but rather access to direct presentational meaning itself.

This approach resolves the Knowledge Argument. It demonstrates that complete physical knowledge does not constitute complete knowledge - not because there exist non-physical facts to be discovered, but because presentational meaning exists in a fundamentally different way that cannot be captured through referential means alone.

Philosophical Zombies

Consider three formulations of the zombie question:

(7) "Is a being physically identical to me but lacking consciousness possible?" This is the traditional zombie question, but it's imprecise as it uses only referential terms.

(8) "Is a being physically identical to me but lacking blackness possible?" This formulation improves upon the first by specifying a particular quale, but still relies entirely on referential terms.

(9) "Is a being physically identical to me but lacking █ possible?" This represents the real question, as it properly incorporates a non-referential term.

As in the other sections, (7) and (8) do not present the actual thought experiment, where as (9) does, so it is (9) that we want an answer to.

The critical insight emerges when we attempt to pose the question: To even specify what the zombie would lack, we must experience █. The very attempt to formulate the zombie possibility requires instantiating what the zombie supposedly lacks. This observation isn't merely a clever semantic trick - it reveals something fundamental about the nature of consciousness and its relationship to reference.

When dealing with purely referential terms like "consciousness" or "blackness," we can coherently imagine their absence. These terms point to experiences without presenting them directly, allowing us to conceive of a being lacking what the terms indicate. However, the situation changes dramatically when we consider non-referential terms. We cannot coherently imagine lacking █ because the very specification of what would be lacking necessarily presents it. The attempt to conceive of its absence paradoxically requires its presence. So it lacks counterfactuals.

The implications are profound: The zombie argument functions effectively within the domain of referential terms but collapses when applied to non-referential terms. Since consciousness fundamentally involves non-referential meaning, as demonstrated throughout this paper, the zombie argument fails to establish its intended conclusion about the nature of consciousness. The purported conceivability of zombies stems from conflating referential descriptions of consciousness with the direct presentational nature of conscious experience itself.

This reformulation through non-referential terms doesn't merely challenge the zombie argument - it dissolves it by revealing that the very attempt to formulate the possibility of zombies requires what zombies are supposed to lack. The zombie argument, like the Knowledge Argument before it, stumbles not on empirical grounds but on the logical structure of how we must engage with non-referential meaning.

The zombie argument was supposed to argue that there are qualia. But here we have already *demonstrated* that there are qualia. So, while the zombie argument itself doesn’t go through, we’ve established a much stronger argument that gives the conclusion that qualia exist.

Spectrum Inversion

Consider three formulations of the spectrum inversion question:

(10) "Could someone's color experiences be systematically inverted relative to mine?" This is the traditional formulation of the spectrum inversion thought experiment, but it's imprecise as it uses only referential terms.

Because this paper outputs in only black and white, we’ll change the question slightly in an obvious way that obviously keeps the point of the Spectrum Inversion question, without loss of generality.

(11) "Could someone's experience of a rectangle be what I experience as a triangle?" This formulation improves upon the first by specifying particular qualia, but still relies entirely on referential terms.

(12) "Could someone's █ be my ▲?" This represents the real question, as it properly incorporates non-referential terms.

As in the other sections, (10) and (11) do not present the actual thought experiment, whereas (12) does, so it is (12) that we want an answer to.

The critical insight emerges when we attempt to pose the question: To even specify what would be inverted, we must experience both █ and ▲. The very attempt to formulate the inversion possibility requires instantiating both qualia in their non-inverted form. This observation isn't merely a semantic trick - it reveals something fundamental about the nature of conscious experience and its relationship to reference.

When dealing with purely referential terms like "rectangleness" and "triangleness," we can coherently imagine their inversion. These terms point to experiences without presenting them directly, allowing us to conceive of them being swapped or rearranged. However, the situation changes dramatically when we consider non-referential terms. We cannot coherently imagine ▮ being ▲ because the very specification of what would be inverted necessarily presents each quale exactly as it is. The attempt to conceive of their inversion paradoxically requires their direct, non-inverted presentation.

The implications are profound: The spectrum inversion argument functions effectively within the domain of referential terms but collapses when applied to non-referential terms. Since color experience fundamentally involves non-referential meaning, as demonstrated throughout this paper, the spectrum inversion argument fails to establish its intended conclusion about the private or subjective nature of conscious experience. The purported conceivability of spectrum inversion stems from conflating referential descriptions of color experiences (or, in our case, shape experiences) with the direct presentational nature of the experiences themselves.

This reformulation through non-referential terms doesn't merely challenge the spectrum inversion argument - it dissolves it by revealing that the very attempt to formulate the possibility of inverted qualia requires presenting each quale exactly as it is, not as inverted. The spectrum inversion argument, like the Knowledge Argument and zombie argument before it, stumbles not on empirical grounds but on the logical structure of how we must engage with non-referential meaning.

The spectrum inversion argument was supposed to demonstrate the private, subjective nature of conscious experience by showing that different people could have systematically inverted color experiences while behaving identically. But here we have demonstrated something stronger: The very nature of non-referential meaning shows that such inversion is not merely empirically false but logically incoherent. The attempt to specify what would be inverted necessarily presents each quale as itself, making the notion of inversion fundamentally contradictory when applied to non-referential terms.

This resolution carries significant implications for our understanding of consciousness and qualia. Rather than supporting the traditional view that conscious experiences are purely private and subjective, the non-referential analysis suggests that qualia have an objective presentational nature that resists both referential reduction and subjective variation. The seeming possibility of spectrum inversion arises only when we mistakenly treat all aspects of consciousness as referential, overlooking the direct presentational nature of conscious experience itself.

However, there is a spectrum of verifiability: verifiability about objects, intersubjective verifiability, and intrasubjective verifiability. It is the latter that we have in (12). Thus, (A) I can verify that my ▮ is not my ▲. But (B) I cannot verify that my ▮ is not your ▲. However, you can verify that your ▮ is not your ▲. The upshot is that the spectrum inversion thought experiment goes through for referential terms and it goes through for non-referential terms (case (B)) and this is consistent given intrasubjecte verifiability (A) and (C).

Conclusion

The introduction of non-referential terms into classic consciousness thought experiments reveals striking patterns and offers novel insights that were previously obscured by purely referential analyses. Across all four thought experiments examined—the Hard Problem, the Knowledge Argument, Philosophical Zombies, and Spectrum Inversion—we observe a consistent transformation when non-referential terms are properly incorporated.

First, in each case, the traditional formulation using only referential terms fails to capture the actual phenomenon under investigation. The progression from purely referential terms (like "consciousness" or "blackness") to mixed referential/non-referential formulations (incorporating █ and ▲) doesn't merely clarify these thought experiments—it fundamentally transforms them. This transformation reveals that many apparent philosophical puzzles arise from conflating referential descriptions of consciousness with the direct presentational nature of conscious experience itself.

Second, we discover a recurring pattern: the attempt to formulate these thought experiments using non-referential terms often requires instantiating the very experiences that are supposedly in question. This creates a kind of self-referential paradox that dissolves rather than solves the original problems. In the zombie case, specifying what zombies lack requires experiencing it; in spectrum inversion, describing what would be inverted necessitates experiencing the non-inverted qualia.

Third, these analyses suggest that consciousness has an inherently presentational aspect that cannot be captured through purely referential means. This isn't merely a limitation of our descriptive capabilities—it points to a fundamental feature of consciousness itself. The non-referential approach demonstrates that certain aspects of conscious experience resist both referential reduction and traditional philosophical analysis.

Perhaps most significantly, this investigation shows that incorporating non-referential terms doesn't merely add new tools to our analytical toolkit—it radically transforms our understanding of consciousness itself. The apparent intractability of these classic thought experiments may stem not from the mysterious nature of consciousness, but from attempting to analyze inherently presentational phenomena using exclusively referential tools.

While this approach resolves certain philosophical puzzles, it simultaneously opens new questions about the nature of non-referential meaning and its role in conscious experience. Future work might explore how this framework applies to other philosophical problems in consciousness studies, or investigate the relationship between referential and non-referential aspects of conscious experience more broadly.

What emerges is a new perspective on consciousness—one that acknowledges both its referential and presentational aspects, and recognizes that some philosophical puzzles about consciousness dissolve when we appropriately account for its non-referential nature. This suggests that progress in understanding consciousness may require not just new answers, but new ways of asking questions that incorporate both referential and non-referential terms.

References

1. Merriam, P. and Habeeb, C. (2024). Introduction to Non-Referential Terms, the Calculus of Qualia, Truth and Meaning (unpublished manuscript cited in the paper)

3. Byrne, A. (2020). "Inverted Qualia." The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/qualia-inverted/
4. Kind, A. (2020). "Qualia." The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/qualia/
5. Nagel, K. (2019). "The Knowledge Argument." The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/knowledge-argument/
6. Kirk, R. (2019). "Zombies." The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/zombies/
7. Chalmers, D.J. (2020). "The Hard Problem of Consciousness." The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/consciousness-hard/