Perspective ~~Dualism~~

The idea that there is something irreconcilably different between mind and brain -- i.e., between our inner mental world and the physical substrate that appears to support it -- has a long and varied history, enduring with considerable popularity to the present day. Descartes’ formulation of *substance dualism* established a clear, but also clearly problematic, version of this idea, postulating that the mind is constituted of an entirely nonphysical substance, somehow independent yet interacting with the physical substrate of the brain. It is this awkward and contradictory relationship of dependence and independence that makes such dualist frameworks so difficult to understand and accept. And yet, Chalmers’ more recent endorsement of a similar form of dualism in the face of these difficulties has been remarkably influential, propelled by compelling arguments about the severe limitations of any attempt to provide a satisfying account of our subjective experience in terms of physical properties of the brain (Chalmers, 1995). Specifically he usefully distinguished the truly *hard problem* of explaining *qualia* -- i.e., “what it feels like”, from the various other easier questions in the broader scope of thinking about the mind, brain, and consciousness. It is this hard problem of qualia where dualism gains its strongest support, and this is the focus of the present paper.

We argue that there is a conceptually simple way of understanding the true nature of this dualism, called *perspective dualism*, that has been discussed in various ways by a number of people over the years, but has perhaps not yet received sufficiently focused attention and elaboration. The core idea is that there is a fundamentally irreconcilable duality between the *subjective, first-person perspective* (looking out from the inside) and the *objective, third-person perspective* (looking in from the outside). The subjective perspective is what a physical human brain feels from the inside looking out: each human being experiences this first-person subjective experience of their own physical brain (and body), and, definitionally, what it feels like to be that brain is what each such person experiences. Only human brains can generate the subjective experience of being a human brain (so far at least), and each such brain supports one and only one unique such subjective experience, because there is a strict identity between any given physical system (e.g., the brain) and whatever subjective perspective might result from the property of *being that system*.

It follows from this identity relationship that achieving an objective, third-person understanding of what it feels like to be a human brain is impossible, thus giving rise to the irreconcilable dualism between subjective and objective perspectives. If I did not already occupy the subjective perspective of a human brain, no amount of verbal explanation, poetry, movies, or other forms of external communication can convey what this subjective experience feels like, because the only way we can truly understand such information is by reference to our own individual subjective experience. Thus, human brains can share their subjective experiences in expressive ways, but those expressions would be meaningless without the primary referent of the subjective experience itself. The degree of actual sharing across two individuals presumably depends on the extent that their brains generate similar subjective experiences, and the quality of the shared communication. However, given the impenetrable barrier of our own individual subjective perspective, it is definitionally impossible to verify with any certainty that a subjective experience has in fact been shared --- all we have is each individual’s subjective sense that some kind of sharing has occurred, but again this remains forever subjective and there is no possibility of a definitive objective verification.

In this way, we can see that there is a conceptually self-consistent framework of perspective dualism, which avoids the difficulties of substance dualism, while retaining the essential truth of Chalmers’ argument that the hard problem requires some kind of dualism: you really can’t explain the nature of subjective experience using objective, third-person terms describing the detailed workings of the brain for example.

In the following, we first articulate a series of axioms that progressively elaborate this self-consistent perspective dualism framework, and establish the foundation upon which some form of objective understanding can emerge, which then allows us to characterize the distinction between the subjective and objective perspectives. To be clear, we argue that objective understanding is strictly speaking impossible, and that all understanding is fundamentally subjective, but at least we can construct some basic ideas that, if accepted by other subjective beings, provide a self-consistent way of talking about the objective universe in relation to our primary subjective experiences. We further elaborate this objective worldview in the context of some relatively well-established principles of brain function in relation to our conscious experience, to provide a self-consistent explanation for some important features of our subjective experience as human brains. Finally, having clearly articulated this perspective dualism framework, we relate it to a range of prior ideas in the literature.

## Axioms of Perspective Dualism

The following assertions or axioms are organized in a progressive manner, such that any particular individual may choose to depart from the progression at any point, consistent with the overarching dominance of the first axiom. Of course, people could choose to believe any random subset of such axioms as well, but we believe that this ordering is logical, and useful for establishing the overall framework, even if at first it may appear unnecessarily pedantic. These axioms provide a basis for then discussing specific neuroscience facts that bear upon the nature of perspective dualism, again in a progressively elaborated manner.

### Foundational Axioms

1. The subjective perspective is primary, and, ultimately, all that can be known with any certainty. This is one of the implications of Descartes’ famous *Cogito, ergo sum* statement: all that we can be certain of is that we are here, thinking. Everything else is a construct that must be established within each individual’s subjective realm.
2. The subjective perspective is unique to each individual. I have my own personal subjective experience, and you have yours, and none of us have managed to directly experience the true first-person subjective experience of another person. This establishes the subjective boundary between the self and everything else.
3. An individual can choose to believe that their subjective experience is the product of an objective reality that exists outside of their own subjective world. Or they could choose not to (i.e., solipsism), including an optional belief in the existence of other beings (which could be construed in other non-objective terms, e.g., as a collection of spirits in a magical world). More generally, we can have all manner of subjective beliefs that can be inconsistent with each other and with those of other people etc.
4. Assuming a belief in an objective reality (the *ontic postulate;* Jilk??), an individual can further choose to believe that the nature of this objective reality can be understood in some kind of systematic manner within their own subjective world of ideas and beliefs. This includes a recognition that there are other people out there in the objective world that are highly similar in many ways to the subjective self, such that it makes sense to regard these others as likely to be an important source of information upon which to build an understanding of objective reality (e.g., the subjective self can take advantage of having an objective external perspective on others to cross-reference against their own subjective self).
5. A specific version of such subjective understanding of the objective world holds that the “true” nature of the objective world is characterized by a set of ideas that are all mutually compatible (i.e., self-consistency), including a consistency between the nature of subjective experience and properties of the objective physical world. (This does not mean that we should expect any kind of automatic consistency between the *contents* of our subjective beliefs and the facts of the objective world -- our objective understanding must accommodate the fact that we can also believe all manner of random things that do not directly correspond to objective facts).
6. The project of developing this self-consistent objective understanding involves a process of attempting to articulate this self-consistent set of ideas, in such a way that other subjective beings who have adopted something like this set of 7 axioms are also willing to endorse the self-consistency of the accumulated set of ideas. This is the project of science, with the reproducibility criterion of the scientific method being the practical mechanism of verifying self-consistency: if everyone can replicate an experimental result across labs and across time, by following a particular experimental method, then we can all admit such a thing as an established “fact” of objective reality (subject of course to subsequent reinterpretation, contextualization, etc).
7. Likewise, if everyone agreeing to this set of 7 axioms can agree about the logical self-consistency of a set of ideas such as those being articulated here, then these ideas also become a foundation for building an objective understanding.

### Basic Neuroscience Facts

Building upon the 7 foundational axioms articulated above, we can now establish a set of basic facts about the relationship between the brain and the subjective experiences of humans having such brains.

1. The ability to think and have a subjective experience that is capable of understanding the ideas presented here depends on specific properties of the human brain. These properties of the brain, and our corresponding subjective experience, develop over the course of our lives, consistent with our accumulated trace of subjective first-person memories. We know of no other entity that shares the kinds of subjective experiences that we each have, outside of those with human brains. The objective evidence supporting this correspondence between the brain and the nature of our subjective experience is extensive, including the subjective effects of objective manipulations such as electrical stimulation, drugs, neurosurgery, neural trauma of all sorts, etc.
2. Setting aside any detailed consideration of the correspondence between subjective experience and objective brain properties for the moment, the basic fact of this correspondence is sufficient to conclude the fundamental premise of perspective dualism: The detailed qualia of my own subjective experience -- what it feels like to be me -- is by definition the subjective perspective of being a human brain. In other words, there is an empirical identity between the objective presence of human brains and reported subjective experiences of others that are sufficiently closely corresponding to those of my own subjective experience, that I can be sufficiently certain that my own brain produces my subjective experience.
3. The objective version of fundamental axiom 2 is that each distinct human brain produces a correspondingly distinct subjective perspective. There is something (which we will discuss in detail below) about the brain that produces a subjective experience from the perspective of being that brain, which we experience as a coherent *self* of some sort, as independent from others.

If we accept points 1-3 above, a number of other logical possibilities and further questions follow. Could other kinds of physical systems produce other kinds of subjective experiences that would be experienced in some way inside such systems? Most obviously, what about the brains of other non-human animals, which objectively resemble ours to varying extents? But more remotely, what about complex systems like my laptop? Or systems at different scales, such as entire societies, the earth, solar system, galaxy, etc? While these are important “practical” questions for understanding the nature of our own subjective experience in a comparative sense, it is important to emphasize that none of that alters the fundamental premise of perspective dualism: all available evidence strongly supports the identity of mind and brain as two perspectives on the same thing. Before we explore these more detailed practical considerations about brain -- mind properties, we elaborate a few logical implications of the framework so far.

### Logical Implications

The following implications can be derived from the basic premises established to this point:

1. While we can (and easily do) conceptualize the subjective self as a kind of physical thing, it is not. It is the subjective perspective of being a physical thing, and cannot actually be separated from that physical thing. This is the dividing line between substance dualism and perspective dualism, and it obviously carries some heavy burdens, consistent with materialistic monism. However, unlike materialism, perspective dualism starts within a foundation of subjective experience as primary, and constructs objective materialism within that foundation, as an ideal, yet ultimately unreachable goal: our understanding of the objective world will always be filtered through our own subjective lens. While it would indeed be nice if the soul was somehow a magic substance independent of the physical world, yet somehow capable of interacting and controlling it, the obvious inconsistencies of this idea are incompatible with an approach founded on self-consistency as the primary guiding principle of establishing truth.
2. If I am identically my own brain, it is logically and physically impossible for me to *also* be *your* physical brain in addition to being my own brain. Thus, our subjective experience is inevitably isolated: it is impossible for any person to directly experience the subjective world of another. This means that there is no way to directly establish objective truth, defined as something that is unequivocally shared across multiple people. We are each separate sovereign subjective domains, and objective reality is a shared, evolving construct among sovereign states that choose to participate in incrementally building such an edifice.
3. This inability to directly share subjective experience is why qualia will forever remain the *hard problem* -- indeed it is the *impossible* problem. Interestingly, many people confront the hard problem with denial: failing to recognize the impenetrability of the subjective fortress, presumably because of the beguiling ease with which we actually can empathize with others in practice. We can also easily articulate many properties that we now know about our brains, and how those seem to correspond with our own subjective experiences, giving a false sense that somehow conveying these objective facts would be capable of delivering the corresponding subjective experience to some other being that otherwise lacks it. It only works if you have the source subjective experience yourself already. And, even there, we truly cannot be certain that any of our personal subjective experiences are shared by others. Both extremes seem equally improbable: that all of our subjective experiences are shared, or that all are unique -- the unknowable truth is likely somewhere in between.

## Coherent Neuroscience of Consciousness

In the same way that we are tempted to think of the self as a physical thing that could somehow be separated from the physical brain, there is a tendency to talk about consciousness as an independent kind of “thing”. It is not. Indeed, *consciousness* is likely to be synonymous with the subjective experience of being a human brain. As such, the entire scope of psychology and neuroscience is relevant for establishing a detailed understanding of why our subjective experience has certain properties. Here, we focus on the few major properties that most people identify as likely to be uniquely human, and most important for differentiating what we informally characterize as “unconscious” from “conscious” states.

Phenomenologically, these special properties center around the sense of awareness of being aware --- the fundamentally “meta” aspect of consciousness. Sometimes we can specifically focus on this metacognitive self-conscious awareness, but even when we might be “lost in our thoughts”, there remains a sense of a coherent “self” acting as a perceiver that stands apart from what is being perceived. Many current neuroscience-based theories of this awareness aspect of consciousness have emphasized the importance of various constructs that seem to share at least a family resemblance to each other, including: *global workspace* (Baars et al; Dehaene et al)*, blackboard* (Mumford, 91), *recurrence* (Lamme, 2006), *coherence and persistence* (Cleeremans et al), and *integrated information* (Tononi, Koch).

In our own framing of these ideas, the core idea is that the human neocortex has extensive bidirectional connectivity across many areas (recurrence; Lamme 2006) that supports a *global constraint satisfaction process* (Hopfield, 1984; Herd et al, 2021) where information across all sensory modalities, motor planning areas at multiple levels and time scales, emotional areas, language areas, episodic memory, etc all engage in a massive interactive “conversation” of neural communication. This massive “plenary” meeting inside your mind is often focused on trying to make sense of the current sensory inputs, to formulate appropriate behavioral responses, but can be diverted to pursue all manner of other random thoughts. The key property of this massively interactive dynamic is that it produces some kind of coordinated, coherent activity patterns across wide areas of the cortex --- this then corresponds to a subjective experience of a *unitary* state of consciousness --- at every moment, there is something specific that we are thinking about or experiencing, which is dominating all other possible things that we otherwise could be thinking about. And then it moves on to the next such thought, etc.

At each step in this flow of conscious states, each brain area is essentially “aware” of what the other brain areas are saying, and this “mutual awareness” of the global workspace / plenum / blackboard state seems like a good fit with the subjective sense of awareness we seek to capture. Dennett has derided this kind of characterization of awareness in terms of a dualistic *Cartesian theater*, where there is an immaterial *self* that is somehow standing apart, watching the contents of subjective experience as if on a movie screen. However, if we instead see this as a subjective-perspective self *emerging* from the mutual-awareness of each part of the brain interacting with the others, perhaps it makes more sense. Much of Dennett’s critique rests on a strident denial of any form of dualism in favor of a strictly materialistic worldview, which is consistent with his denial of the validity of Chalmers’ formulation of the hard problem. However, if you begin instead with the acceptance of the first-person view as a perfectly valid and distinct perspective, then perhaps the Cartesian theater instead becomes another apt way of capturing the unique subjective perspective of a massively bidirectionally connected cortical system.

The neocortex, which is most expanded in the human brain relative to other mammals, and is expanded or only present in mammals relative to other animals, is unique in supporting the all-important bidirectional connectivity at the core of the above account. Furthermore, the human brain has an extensive hierarchical organization of higher-level brain areas that are relatively far away from sensory inputs in terms of network connectivity, which enables these areas to establish more enduring coherent activity states that persist over time in the face (or absence) of sensory inputs. The pinnacle of these higher-order areas is the prefrontal cortex, which is particularly expanded in humans, and widely implicated in metacognitive awareness, cognitive control, and more abstract levels of thought. Thus, these architectural aspects of the cortex are likely to play important roles in the relatively enduring coherence across time of our conscious thought process, relative to other animals which may live much more “in the moment” driven by immediate sensory inputs.

## Comparative Consciousness

The foregoing points provide a basis for revisiting the important issues of comparative consciousness raised earlier: if we can identify our subjective experience with the physical substrate of the human brain, what kinds of subjective experiences might other physical systems support? What really makes the brain so special among everything else in the known universe, aside from the critical fact that it happens to be the thing engaging in this process of trying to understand its own nature? Inevitably, this exercise is colored by the particulars of our own brains, and limited by the scope of our imaginations. Furthermore, because we can never actually subjectively experience anything other than our own first-person perspective as a human brain, all of this remains pure speculation. Nevertheless, perhaps it provides some exploration of at least the local space of possibilities that can be extrapolated around the case of the human brain, affording a somewhat more systematic sense of the boundaries and properties of this space.

With these caveats in mind, the following list can be defended as a minimal set of properties that a physical system should have in order to generate a subjective experience that we could recognize as having some essential features in common with our own. The main focus here is on attempting to capture what a system might need to be able to distinguish its own *self* from its surrounding environment, as a necessary prerequisite for having a meaningful subjective perspective that could be identified as such (i.e., a *self-perspective*).

1. *Physical coherence and identity*: At a basic physical level, it is reasonable to assume that any kind of self-perspective would require a stable, physically contiguous, interacting system of some form --- otherwise the subjective states would be transitory and lack spatiotemporal coherence of a form that seems necessary to establish a “thing” that could possibly be differentiated from any other random collection of matter and energy. Specifically, a self-perspective requires having a well-defined and unique trajectory through space and time.
2. *Internal complexity*: The richness of subjective experience depends on a physical system capable of exhibiting many meaningfully different states, in an information-processing sense. It is clear in the case of our own brains, and other information processing systems such as computers, that the ability to exhibit many different states that each change the functional behavior of the system is an index of the total memory and processing capacity of the system (e.g., the number of synapses in the brain or the number of transistors in a computer). The integrated information theory (Tononi etc) provides an example of an attempt to define consciousness in terms of objective information-based parameters of this sort. However, while such minimalist constraints do exclude many possible physical systems, they may not usefully exclude a large class of things that do not seem likely to have self-perspectives (cite: Scott Aronson blog), so they may not be sufficient on their own.
3. *Self-control / agency*: A physical system that cannot act in the world and control itself in some meaningful way would seem to lack an essential ingredient for self-perspective. If a physical system is completely at the whim of external forces, then it is not clear that its subjective experience would be differentiated from these external forces. In other words, to define the concept of a self that is distinct from its environment, that self-embodying thing needs to be able to act independently to at least some extent. There is a famous example of a sea squirt that eats its brain once it settles down in a permanent location on a coral reef: the brain is required to control action, and conversely, in the absence of action, no coordinated central control system is required.
4. *Self-model and self-awareness*: A self-perspective requires some ability to represent its own state, and perhaps a broader sense of itself as a distinctive, individuated system. It may not require a full-blown metacognitive capacity, and full self-recognition abilities (e.g., the ability to recognize itself in a mirror), but some minimal ability to internally represent the extent and current state of the system comprising the physical basis of the subjective experience seems like an important element of any self-perspective supporting system. This ability may also be essential for any form of effective self-control: effective action requires knowledge of the current state of effectors, for example.
5. *Perception*: The ability to perceive distinct states of the external environment is likewise probably necessary for any meaningful form of self-control, and almost certainly is important for giving a system a sense of its own subjective situation within the external world. It is also likely essential for differentiating the self from the external environment.

These criteria exclude many different physical systems from being plausible candidates for supporting self-perspective level subjective experiences, while also not being overly restrictive in terms of possible artificial systems that might support such subjective experiences. In particular, there is nothing here that requires the system to be biological, or to have a specific type of perceptual and control system (i.e., brain). But it does exclude things like rocks, which inevitably plague the minds of philosophers contemplating the nature of subjective experience. More pressingly, it excludes things like laptops, which pass the first 2 criteria, and represent a particular challenge for integrated information theory (cite Aronson), but fail on the remaining criteria that provide a basis for really distinguishing a sense of the subjective self from the surrounding environment.

A particularly interesting test case is Twitter, which has a bidirectional conversation dynamic that is perhaps quite similar to the plenary space in the human neocortex. There are trending topics that function much like the current focus of conscious awareness, dominating the conversation for a while before passing inevitably on to other topics, etc. However, Twitter also lacks a coherent spatiotemporal existence and any kind of self-control / self-model and perceptual abilities -- it isn’t really a distinguishable, coherent, physical “thing” acting in a world. It is just the conversation. Similar considerations apply to large-scale systems like the earth etc.

In the following, we consider the specific cases of animals and AI’s in reference to the above criteria.

### Animal Subjective Experience

Pretty much any animal, even a single-celled organism, can meet at least a minimal version of the above criteria. Thus, the relevant question here becomes the degree to which the various animals might experience something more similar to our own subjective consciousness. The simple answer here is inevitably that animals having brains that are most similar to our own are likely to have the most similar subjective experiences. In particular, the importance of bidirectional connectivity and hierarchical brain organization in the human brain provide a useful metric that generally tends to align well with our intuitive understanding of the relative richness of the inner life of different species (Figure X).

|  |
| --- |
|  |
| Figure X: The left panel shows brain size as a function of body size (both measured by weight), where those above the green line have relatively more neurons per body size, and thus could be considered potentially more likely to have richer mental lives. This measure generally agrees with our overall intuitions, although dolphins are remarkably close to humans on this measure. The right panel shows the number of neurons in the neocortex (and bird equivalent, pallium) relative to total numbers of neurons in the brain for a selected set of species organized into different clades: *artiodactyls* are even-toed ungulates such as pigs, antelopes, deer, goats, and cattle, while *afrotherians* include elephants, aardvarks and golden moles, and *eulipotyhphlans* include hedgehogs, shrews and moles. By cross-referencing with brain mass, it is clear that neuron density varies significantly. In terms of likelihood of strong recurrent processing, the number of neurons in the neocortex is the relevant quantity. Humans have the most cortical neurons (and also a correspondingly large number of cerebellar neurons, not shown), but elephants still have quite a lot, while parrots and related birds really pack a lot of neurons into their small brains. |

Another critical practical consideration is that the ability to communicate certainly helps in evaluating the nature of subjective experience of others: if another being is telling us in great detail about their subjective experiences, we can get a much better sense of what that experience is like. In this regard, it is interesting that great apes that have been taught various means of communication exhibit a significantly different profile of communication than even young human beings first learning language. Specifically, much of the language output is an endless “discussion” of food and other obsessions, with relatively rare examples of anything more “thoughtful” or “reflective”. Thus, perhaps some particular aspects of the human motivational orientation toward complex social interaction is also important for enabling the development of complex cultural systems that provide sufficient education and foundation for more complex forms of thought (Thomasello). Again, there is the persistent sense that other animals are much more bound to the immediate moment than people, and thus their subjective experience is less unique, individual, and subject to self-control and self-determination.

### AI Subjective Experience

Interestingly, while the self-perspective criteria are broadly inclusive of animals, they are not so obviously inclusive of many current forms of AI. Most existing AI models are not embodied in any kind of physical system, robot or otherwise (even a purely virtual embodiment within a virtual environment), that would then support the ability to distinguish a self from the surrounding environment. Furthermore, even among systems that are embodied, they are often controlled by external means in one form or another, and don’t really end up taking full control over their own physical selves in the way that all animals do. For example, many agent-based AI systems are trained through externally-provided scalar reward signals, and their learning is a monotonic function of such signals, without any kind of rich encoding of the internal state of the system, or a full-fledged self-control system that is truly capable of autonomous decision-making.

In addition, almost all of the current AI models use a feedforward form of activation propagation that is highly efficient computationally, especially in supporting error backpropagation learning, which requires directed, not cyclic computational graphs. When recurrent networks are included, they are typically of a highly constrained form. These properties are very different from the broad bidirectional connectivity and corresponding constraint-satisfaction processing that is likely central to the nature of our own subjective experience as discussed above. As a result, it is not likely that these models are supporting a rich, complex, inner subjective experience, consistent with our overall sense of their observable behavior. Perhaps by incorporating more of the properties thought to be important for human conscious experience, future models will become more like us in their subjective experiences, at which point all manner of interesting moral issues will need to be addressed.

## Summary

In summary, the perspective dualism framework has the following advantages:

* It embraces the fundamental dualism emphasized by Chalmers’ framing of the hard problem of consciousness.
* It likewise avoids a simplistic denial of this dualism that a purely materialistic framework requires: we strongly believe that it really is not possible to explain what it feels like to be a human brain in purely objective terms.
* It avoids the logical inconsistencies of substance or property dualism, and does not require a nonphysical substance or property that still interacts with the physical world, while still establishing a very strong form of dualism.
* It recognizes the primacy of subjective experience, per the central insight of Descartes’ *Cogito ergo sum,* thereby accommodating all the severe epistemological and practical barriers to achieving an objective understanding of the world.

Despite these advantages, there may remain a persistent concern: have we really explained why it feels like something to be a human, and why it feels the specific way it does? Indeed, the whole point here is that this would require the ability to provide an objective explanation for subjective experience, which we argue is strictly impossible. So, like many other possible questions that can be easily formulated, these questions may just not be answerable. Except to accept that your own subjective experience is exactly what it feels like to be you, and that’s a sufficient explanation for each of us.

# Comparison with Other Frameworks

* Russelian monism
* Nagel: perhaps the closest case I know of..
* Others?