Undergraduate Philosophy Journal of Australasia www.upja.online
Vol. 5.2, February 2024

Ned Block's homunculi-headed robot and functionalism

Jack Blackman[1]
Victoria University of Wellington

Abstract

Ned Block posed his thought experiment of the homunculi-headed robot in his paper '*Troubles with Functionalism*' to try to defeat functionalism, a leading theory within the philosophy of mind, which concerns the nature of mental states. The robot was meant to defeat functionalism by showing how functionalism attributes mental states inappropriately, as beings such as the robot would have had mental states under functionalism, despite possessing no qualia. Block's argument rests upon two incorrect assumptions of qualia that this paper exposes as being incorrect. Firstly, Block presupposes that due to qualia being an innate part of human beings, it cannot be accounted for under functionalism. Secondly, Block applies too narrow a perspective onto what qualia can be, and where and how it can subsist, to be considered valid. I attack these two incorrect presuppositions to exhibit how Block's robot cannot be considered to have defeated functionalism.

1. Introduction

Ned Block's homunculi-headed robot is a thought experiment that was posed to exhibit how functionalism is unreliable in posing a paradigm that explains mental

[1] Jack is a student at Te Herenga Waka, Victoria University of Wellington, studying towards a law degree, and a bachelor of arts degree, majoring in philosophy. He is primarily interested in philosophy of mind, AI, and phenomenology.

states and their relationship to consciousness. The purpose of Block's robot was to highlight how functionalism was too liberal in attributing mental states to beings. Block's robot has the same physical appearance as a human, and similarly exhibits the same behaviours as a human. However, it replaces the human brain with a multitude of little people who are tasked with executing certain operations given certain inputs, which mechanically compels the robot to behave in certain ways. This thought experiment was designed to embarrass the functionalist account of mental states, exposing how they do not accurately explain how mental states engender consciousness. This paper disagrees with the conclusion Block's robot comes to, due to a perceived mistake in how Block perceives the nature of qualia, and his imaginative restrictions on what qualia can be. To suffice this thesis, this paper is split into three distinct sections, which will be provided below, for the sake of clarity:

1. Functionalism: What it is, how we can come to understand it through real life accounts, and how Block's robot relates to it.
2. The innate nature of qualia: What qualia is, what it means for it to be innate, how this relates to the theory of functionalism, and a thought experiment to exhibit how functionalism can possess innate components.
3. Imaginative restrictions on qualia: How our own conscious bias affects our construction on qualia, the wiring of our brains, and the mistake made in common inferences of consciousness.

By the conclusion of this paper, it should be evident that Block's homunculi-headed robot does not defeat functionalism.

2. Functionalism

Functionalism is a well-established theory within the discipline of philosophy of mind, which focuses on the nature of mental states. It is functionalism that Ned Block attempted to confute with his robot. An example of a mental state is possessing a belief, having a desire, being in pain, etc. Functionalism asserts that "what makes something a mental state of a particular type does not depend on its internal constitution, but rather on the way it functions, or the role it plays, in the

system of which it is a part".[2] Functionalism is like a tap; a tap is a tap regardless of what it is made out of, so long as it performs its proper function of dispensing water. Hence, mental states are "fundamentally relational", characterised by relationships with things such as "stimuli, output behaviours, and other mental states".[3] Before the paper proceeds, it is vital that three concepts; consciousness, force, and states, are defined, as they are incredibly relevant as the paper progresses. These will be provided below:

Consciousness. The qualitative experience a being possesses, granted they have a cognitive system consisting of distinctive states which act on and influence each other.

Force. The essence of a particular state.

States. The realisation of particular conditions which serve independent functions yet are also influenced by other states acting within the same cognitive system of consciousness.

Furthermore, this paper will assume the truth of functionalism, and its tenets as pertinent to this essay. As Block set out to destroy functionalism, and this paper critiques Block, the sake of simplicity demands that this paper assume the truth of functionalist theories.

2.1 Pain and functionalism

To root the theory of functionalism in reality, imagine you were experiencing the sensation of pain. Under functionalism, it is not enough for a being to simply experience the sensation of pain, for such a state to be considered a state. It must experience the state of pain through a system of causal relations, by which the experience of pain is but a constituent within it. Therefore, pain can be characterised simultaneously with other internal states, such as nerves, behavioural outputs, and emotional responses to physical reactions. All of this takes place within a causal

[2] Levin, Janet (2004) 'Functionalism', *Stanford Encyclopedia of Philosophy*.
[32] Kobes, Bernard (2007) 'Functionalist Theories of Consciousness' in *The Oxford Companion to Consciousness.* Oxford.

network, where the aforementioned internal states are linked to each other and the inputs (damage to the body) and outputs (avoidance behaviours), are abstracted from its material realisation.

2.2 Block's homunculi-headed robot

Philosopher Ned Block's paper *Troubles with Functionalism* is often cited as one of the most influential objections towards functionalism.[4] Block proposed the homunculi-headed robot thought experiment as a *reductio ad absurdum* argument to exhibit how functionalism unrealistically accounts for mental states, through an absence of qualia. Block implores the reader to imagine a figure externally analogous to a human being, yet internally differing vastly in nature from humans due to a radically dissimilar constitution (this figure will from hereon be referred to as the robot).[5] Rather than possessing organs and a nervous system, this body contains a multitude of little people, each tasked with executing a specific action given a specific input.[6]

2.2.1 Interacting with the robot

Suppose that you encountered this robot and asked it how their day was. The robot smiles, telling you that it had a good day, and then begins to inquire into how your day was. This interaction by all societal conventions will be considered perfectly orthodox, yet what's truly absurd about this interaction is that the robot was not acting with an independent whim. Instead, the robot had a society of little people pressing specific buttons which led the robot to perform specific activities. These buttons were illuminated when certain external cues were prompted.

2.3 The robot, and functionalism

---

[4] Block, Ned (1978) 'Troubles with Functionalism', *Perception and cognition issues in the foundations of psychology*, **9**: 261-325.
[5] Block, 278.
[6] Block, 278.

In this thought experiment, the robot can perform a task just as any normal person can. However, under the functionalist doctrine, the robot will be said to have mental states, as the states partake in a role within an internal constitution. When analysing the composition of these supposed mental states, Block's robot is meant to expose the absurd conclusions that functionalism entails. Block stresses that there is no "independent reason to believe in the mentality of the homunculi-head".[7] If the robot cannot feel pain, or experience qualia, then the assertion that "the nature of qualia is to be found in its functional role is erroneous, and functionalism is false".[8]

2.4 Functionalism and consciousness

This paper asserts that Block's argument is that functionalism is guilty of liberalism in assigning mental states, and hence, consciousness, to beings. If Block's argument is true, functionalism permits situations where there is "functional equivalence without qualitative equivalence", meaning that qualia "escapes functional explanation".[9] This collapses the structure of functionalism, by refuting the theory that mental states can be understood by virtue of their collective organisation, rather than their internal constitution. This paper takes a leap forward from here, suggesting that Block purports to restrict the scope of consciousness, by advocating for an understanding of qualia as an innate force. If qualia is an innate force, it cannot be accounted for under functionalism, meaning functionalism cannot be true. This paper will defeat this view.

3. Qualia

This section of the essay will explain qualia in detail and exhibit how it plays a fundamental role in the tension between Block's robot and functionalism. Qualia (quale singular) is a term used to denote states of subjective, conscious experience.

---

[7] Block, 456.
[8] Walsh, Jackie. (2017) 'Can a Functionalist Account for Qualia', *Oxford Philosophy Society.*

[9] Kind, Amy (n.d.) 'Qualia', *Internet Encyclopedia of Philosophy.*

At any given point, we are experiencing a multitude of sensations; we see letters on the screen, feel our body pressed down against a chair, perceive the colour yellow, feel the emotion of happiness. In each of these cases, we are "the subject of a mental state with a very distinctive subjective character".[10] There is some special phenomenological character that defines what it is *like* to undergo these aforementioned states. Broadly, qualia pertains to the quality of our human existence that is accessible solely by virtue of introspection, and experience. Qualia is an incredibly important concept to consider germane to the robot. Whilst with little to no difficulty we can imagine the functional operation of the robot, it is incredibly difficult to believe that the robot experiences qualia. Block's objection is established upon the proposition that whilst the functional replication of the robot as a human can be achieved, it lacks the innate qualia that humans possess to qualify as being conscious.

3.1 Qualia as an innate force

As already mentioned, I posit the tension between Block's robot and functionalism to be rooted in the innateness of qualia. But what does this mean? Block's robot seems to presuppose that qualia is a force which can only be bestowed upon someone innately; that is, they are either born with it, or they don't have it. Because of this nebulous nature, no one has been able to exactly explain how, and why, qualia exists, engendering what is known as the 'hard problem of consciousness'. This problem pertains to "how physical processes in the brain give rise to subjective experience".[11] Qualia as a notion describes the character of what it means to exist. It is a vital component of the human experience, and it is something that Block seemingly cannot envision being artificially replicated. As the robot is able to functionally replicate a human, yet it has different internal wiring, Block is unable to attribute qualia to the robot. This position arises out of a belief that the existence of qualia is determined upon the inner wiring of someone/something. As humans possess the necessary wiring to engender qualia (as it is an innate quality of being human), Block

[10] Tye, Michael (2021) 'Qualia', *Stanford Encyclopedia of Philosophy*.
[11] Berent, Iris (2023) 'The "Hard Problem of Consciousness" Arises from Human Psychology', *Open Mind: Discoveries in Cognitive Science*, **7**: 564-587. https://doi.org/10.1162/opmi_a_00094.

has no problem attributing qualia to humans. However, as the robot contains different wiring, despite producing the same output, Block is unable to attribute qualia to the robot.

3.2 The innate nature of qualia and functionalism

If we were to not presuppose qualia to be an innate quality of humans, then the robot would not pose any threat to functionalism. Otherwise, the robot's success in its functional replication of a human would logically entail its success in replicating qualia. The general presumption that Block articulates his argument upon is that qualia exists within humans by virtue of us being human. This is incongruent with the theory of functionalism, as under a functionalist model, all states are understood and defined through their causal relations to other states. If qualia cannot be functionally replicated, its existence as a state is not contingent on contemporary states. A state that was purely innate could not be accounted for by functionalism, as it would not be considered a state due to it not having any relation with other states. This argument hence follows that qualia and functionalism cannot co-exist. I shall lay out the linear reasoning in favour of this thought below:

1. Any conscious being must possess qualia,
2. Qualia is an innate force which does not subsist upon contemporary states,
3. Functionalism attempts to ascertain the nature of states by asserting that states can be understood through their causal relations with other states,
4. Qualia cannot be a part of a functionalist model as it is innate, and does not subsist nor rely on contemporary states,
5. Hence, functionalism cannot account for qualia, and therefore cannot explain the subject of consciousness.

3.2.1 Innateness in functionalism

If it was shown that an innate force could exist within a functionalist framework, then the previous argument would fall apart at the 4th consideration. This paper argues for the standpoint that a state which possesses an innate force can be understood through either a functionalist lens *or* a singular lens. This is achieved by focusing narrowly on the innate force and will be explicated below.

3.3 Instant coffee

This section of the essay will use the example of instant coffee to offer a compelling argument that an innate force can exist within a functionalist paradigm, or a singular model. To exhibit this, consider a teaspoon of instant coffee. Upon observing this coffee, you will notice a physical, tangible substance which can be held and felt. However, if this coffee was to be submerged in boiling water for 10 seconds, then raised out of the boiling water, what would remain instead would be a teaspoon of boiling water infused with caffeine. Coffee possesses the innate disposition to dissolve in boiling water. However, this disposition is not reliant on contingent forces to *exist*. Regardless of the presence of boiling water, the observed grounds of instant coffee will always possess the disposition to dissolve in boiling water. We can infer from this line of reasoning that the possession of an innate disposition is not contingent on relations to external subjects. A subject can possess an innate disposition through mere virtue of existing in isolation to contingent forces, whether they are activated or not. These dispositions can exist idly, or through the specific activation of their quality, i.e., the coffee ground possesses the ability to dissolve in boiling water. Regardless of the *state* of the disposition, the force still exists nonetheless. This line of reasoning is applicable to the theory of functionalism. Particular states can possess innate forces which don't subsist through causal relations but are merely *realised* through causal relations.

3.4 Objections to the coffee objection

There are two possible objections that I can anticipate to the aforesaid coffee objection. These are the analogy as fact fallacy and an argument from the laws of

nature. This paragraph will discuss these two anticipated objections separately and discuss why they should be dismissed.

3.4.1 Analogy as fact fallacy

The analogy as fact fallacy could be argued against the instant coffee objection. This is because of the possible perceived mistake in suggesting that qualia and coffee grounds are comparable enough to apply similar principles from each other. As qualia directly concerns first-person, perspectival experience, whereas coffee grounds are simply a substance that undergoes no inferred experience, the principle may seem weak to analogise. In response to this, it is vital to remember that the coffee objection simply proves that something can have an innate disposition that is only realised through relations to other substances, but nonetheless, will exist in the absence of such substances. This principle is directly relevant to qualia existing within a functionalist paradigm. Assuming the presupposition that qualia is innately bestowed upon us, it can still be understood through its relations to contemporary states. The coffee objection does not seek to assert that coffee and qualia are tantamount. What it does assert is that the relationships their dispositions have with other substances can be analogised, to exhibit how an innate disposition can exist both independently, and through other substances.

3.4.2 The laws of nature

Furthermore, there is an argument that the disposition of coffee can be explained through the laws of nature. As it can be explained by something beyond itself, this would mean that it does not have an innate disposition, as it subsists through causal relations, rather than solely being perceived through them. This would contravene the nature of qualia (which is merely realised through other substances) and render the analogy invalid. In doing so, it would affirm Block's argument, showing that qualia are an innate quality that cannot be accounted for under functionalism, hence defeating functionalism. The most logical way to evaluate which objection is stronger; that of the coffee grain, or the argument centred around the laws of nature,

seems to be ascertaining whether there is a theoretical arbiter which can identify when something can be considered as innate. This bar will then assert the feasibility of the coffee argument as a valid argument against Block's objection.

3.5 The requisite for innateness

An arbiter of innateness would be a force that can be considered innate within the parameters of the word as imposed in this paper. Once there is an arbiter of innateness, it can be related to qualia, to answer questions about qualia. In finding this arbiter, it intuitively made sense to conduct an inquiry into whether electrons are innate. This is because they are one of the smallest constituents of the universe and are ubiquitous throughout our world. If anything made of matter in our reality was to be considered innate, it would be electrons. Electrons encircle the nucleus of atoms at specific energy levels to influence the character of atoms. Due to this, they are incredibly small in matter, dwarfing that of an atom, meaning it's reliable to utilise as an arbiter for innateness.

3.5.1 The innate quality of an electron

Electrons have rules, dispositions and qualities that constitute what it is to be an electron. Electrons must behave as electrons, as they are in strict obedience to the parameters imposed on them by nature. They underpin everything in the universe, and are an incredibly small force, which cannot be reduced any further beyond themselves. The existence of electrons, and the properties that they occupy, cannot be explained by anything but themselves. They can be perceived and understood through their causal relations to other forces, but they do not subsist upon the activations of these forces for their innate dispositions to exist. For the purposes of this paper, electrons must be considered to be innate. In spite of all this, electrons can be accounted for under functionalist forces.

3.5.2 Qualia and electrons as innate forces

Qualia existing within conscious beings, the notion presupposed by Block which is pivotal to his functionalist attack, is tantamount to electrons existing within the functionalist order of the universe. As electrons and coffee grains can be functionally accounted for within the parameters imposed in this paper, qualia must be treated similarly, ultimately evincing how the innate force of qualia can be accounted for within a functionalist framework.

4. Narrow construction of qualia

I asserted earlier in this paper that Block purports to restrict the scope of qualia, hindering it to a specific organisation which Block appears to maintain as necessary for having mental states, i.e., consciousness. The robot's supposed lack of qualia stems from it not having a brain. It looks the same as us, acts the same as us, and operates the same as us, yet has a queer internal constitution which seemingly prevents it from possessing consciousness. Directly or indirectly, Block asserts that the acquisition of qualia is restricted to those entities with a brain. This discriminates against any other being which could exemplify all the modalities that would lead us to register that being as conscious yet disabling it from being regarded as conscious due to it possessing different internal wiring.

4.1 The necessary wiring

Block's philosophy can be seen to maintain that humans must possess the correct constitution to elicit consciousness. However, when we detach our constitution from how normalised we have learned it to be, it appears to be just as absurd as the robots. Consider this excerpt explaining the present moment from MIT's school of engineering -

> When you read these words, for example, the photons associated with the patterns of the letters hit your retina, and their energy triggers an electrical signal in the light-detecting cells there. That electrical signal propagates like a wave along the long threads called axons that are part of the connections between neurons. When the signal reaches the end of an axon, it causes the

> release of chemical neurotransmitters into the synapse, a chemical junction between the axon tip and target neurons. A target neuron responds with its own electrical signal, which, in turn, spreads to other neurons. Within a few hundred milliseconds, the signal has spread to billions of neurons in several dozen interconnected areas of your brain and you have perceived these words.[12]

Our consciousness is derivative from processes completely tantamount to the homunculi-headed robot, spare for the instruments utilised to necessitate the functions. We are walking homunculi-headed robots assigned with consciousness, however we cannot even be sure in stating that we really do possess consciousness. While I won't stray from the motive of the paper, it's worth noting that a myriad of theories pertaining to this, such as the simulation theory, solipsism, and Pyrrhonian scepticism, carry valid weight and too trouble Block's theory.

4.2 Reasons to believe in qualia

In reference to his scepticism pertaining to the homunculi-headed robot, Block states that "there is no independent reason to believe in the mentality of the homunculi-head, and I know of no way of explaining away the absurdity of the conclusion that it has mentality"[13]. This must be taken in consideration with Block's acceptance of humans possessing consciousness, as due to us having brains, there is an independent reason to accept our consciousness.

4.2.1 Does everyone have brains?

It is commonly assumed that we all have brains and are conscious, but either of these assumptions could be wrong. This is because attributing consciousness is ultimately provisional without access to the other person's brain. Consider Brian Keeley's distillation of this situation -

---

[12] Dougherty, Elizabeth (2011) *What are thoughts made of?* MIT School of Engineering.
[13] Block, 456.

> I happen to think, say, that Ted Cruz would make a great US President and I support his candidacy. Of course, one must be a native-born US Citizen to be eligible for that office. I've never seen his birth certificate. So, I continue to believe he would be a great POTUS, but if somebody provided me with evidence that he was, in fact, born in Scotland, then I'd of course revise that opinion. But part of my reason for thinking he'd be a great President is that I think he's U.S. born. But that reason is defeasible. In the same way, my belief that you have a brain is defeasible.[14]

4.3 Phenomenological traits of consciousness

In troubles with functionalism, Block references Nagel, who states that conscious experience exists "if and only if there is something that it is like to *be* that organism"[15]. As much as we can try to ascertain what beings besides humans have consciousness, it is ultimately futile as no amount of physical inference will enable us to conceive what foreign phenomena may be like. While we can postulate that a pigeon is a conscious being, there is no way that we can ultimately be sure of it. Hence, physical constitution can only take you so far in determining a being's consciousness, due to the lack of diverse qualitative experience that we can recount. Due to this, we do not have the sufficient tools to ascertain what it's like to be a homunculi-headed robot. Block contradicts himself in referencing Nagel's philosophy as he cannot provide an accurate description of what it is like to be that robot, vic. he cannot determine whether or not it has qualia. This suspends Block's case in a philosophical impasse. His inability to assert with certainty as to whether or not there is something that it is like for the robot to be a robot, leaves his case unconditionally incapable of proving itself correct.

4.4 Imaginative restrictions on qualia

---

[14] Bowen, Connor (2019) *Global Consciousness: A Functionalist Neurophilosophical Perspective*. Claremont McKenna College.

[15] Nagel, Thomas (1974) 'What is it like to be a bat?', *The Philosophical Review*, **83**(4): 436.

This is because our conception of consciousness is limited to what we are capable of imagining within the scope of our own consciousness. Because of this, we are fundamentally limited in having a broad perception of what different mental states would be like, and what different conditions of consciousness these mental states could engender. While we can postulate, for example, that inanimate objects are incapable of experiencing qualia, there is nothing we can raise to support this affirmation beyond our blind intuition. Our inability to completely understand our own consciousness, our own qualia, and our own mental states, extends to our inability to epistemically and metaphysically access the existence of these aforementioned states in beings and forms different to us. This bias is why most people, including Block, attribute with certainty that those like us are conscious. Those, like us, who are able to walk, talk, think, express emotion, passion, interest etc. We assume that once organisational prerequisites for consciousness are met, then consciousness shall arise. However, this organisational prerequisite shouldn't be hindered solely to the physical organisation of humans, as we simply cannot purport to maintain consciousness as a phenomena solely experienced by humans.

4.5 Block's restricted imagination

This prior paragraph is essential to understanding the error underpinning Block's philosophy on the robot. To him, it is absurd that the robot be considered conscious, as it is merely a mechanical replication of a human, and therefore the consciousness that accompanies human experience cannot be replicated. Block is unable to imagine that qualia can exist in a way that differs from human qualia.

4.5.1 Human qualia compared to the robots

This is because human qualia is understood as the norm. When you make a witty remark and somebody laughs, you assume to understand the qualitative experience those around you went through when they laughed. However, this assumption is rooted within a fallacy that all humans experience qualia the same. As we only have qualitative, first person access to anything that we personally go through, we are

involuntarily suspended in an unfortunate state of incessant scepticism, where we can only infer that others have similar qualitative experience as us, with no robust means of making sure of this.

4.5.2 A joke in a room full of people

To exhibit this in reality, suppose that you made a joke in a room filled with three strangers, and they all laughed. Your assumption of their qualitative experience will be that, like you, they found what you said humorous, felt an impulse to express this physically to you, leading to the appropriate behavioural output to reflect that. However, you later find out that in that room, one of them had schizophrenia, one of them was high on LSD, and the other person experienced inverted qualia. Your assumption of what they felt, thought, and perceived would be terribly inaccurate, as you have no means to assert what their qualitative experience truly was. Simply because someone has a brain does not assert that their conscious experience is analogous to yours. Everybody that we encounter in life and regard as conscious, we do so because they look like us. Without knowing if they are real, if they have a brain, or if they experience life in a sense similar to us, we maintain them as conscious, due to the similarities we share in physiognomy and behaviour. If consciousness could be confidently inferred merely by virtue of having a brain, as Block supposes it to be, all the idiosyncrasies and syndromes of the brain would reflect replicative experiences of life. Block's ideology would anoint consciousness to someone in a coma. Block's thought suffers from a lack of imagination of what it is like for the robot to be a robot. If we treated the organisational setup of a human organisational as Block does the homunculi-headed robot, we'd observe the infinite differences in the organisation of our brain to anyone else's brain. We'd see the utter lack of understanding of how our brain, mental states, qualia, and consciousness works. It would look as absurd to call us conscious, as it is to call the homunculi-headed robot conscious. This imagination we project onto assuming we are conscious, can be extended for us to conceive the homunculi-headed robot conscious.

5. Conclusion

This paper explored Block's homunculi-headed robot and how it ultimately fails as a rejection for functionalism. The subject of qualia was explored in depth to suffice this thesis. Block's mistake in presupposing that due to qualia being innate it is unable to exist within a functionalist system was exposed and proven wrong. Furthermore, phenomenological exploration of the subject of qualia and how our own experiences trap us in a perspective bias buttressed the fact that Block is unable to prove that the robot does not have qualia and hence cannot object to functionalism. It is with confidence that we can assert that Block's homunculi-headed robot does not defeat functionalism.

References

Berent, Iris (2023) 'The "Hard Problem of Consciousness" Arises from Human Psychology', *Open Mind: Discoveries in Cognitive Science*, **7**: 564-587. https://doi.org/10.1162/opmi_a_00094

Block, Ned (1978) 'Troubles with Functionalism', *Perception and cognition issues in the foundations of psychology*, **9**: 261-325.

Bowen, Connor (2019) *Global Consciousness: A Functionalist Neurophilosophical Perspective*. Claremont McKenna College.

Dougherty, Elizabeth (2011) *What are thoughts made of?* MIT School of Engineering.

Kind, Amy (n.d.) 'Qualia', *Internet Encyclopedia of Philosophy.*

Kobes, Bernard (2007) 'Functionalist Theories of Consciousness' in *The Oxford Companion to Consciousness.* Oxford.

Levin, Janet (2004) 'Functionalism', *Stanford Encyclopedia of Philosophy*.

Tye, Michael (2021) 'Qualia', *Stanford Encyclopedia of Philosophy*.

Nagel, Thomas (1974) 'What is it like to be a bat?', *The Philosophical Review*, **83**(4): 436.

Walsh, Jackie. (2017) 'Can a Functionalist Account for Qualia', *Oxford Philosophy Society.*