Consciousness and Conscious Machines: What's At Stake?

Damien Patrick Williams [0000-0001-6652-2010]

Virginia Polytechnic Institute and State University
Damienw7@vt.edu

Abstract. This paper explores the moral, epistemological, and legal implications of multiple different definitions and formulations of human and nonhuman consciousness. Drawing upon research from race, gender, and disability studies, including the phenomenological basis for knowledge and claims to consciousness, I discuss the history of the struggles for personhood among different groups of humans, as well as nonhuman animals, and systems. In exploring the history of personhood struggles, we have a precedent for how engagements and recognition of conscious machines are likely to progress, and, more importantly, a roadmap of pitfalls to avoid. When dealing with questions of consciousness and personhood, we are ultimately dealing with questions of power and oppression as well as knowledge and ontological status—questions which require a situated and relational understanding of the stakeholders involved. To that end, I conclude with a call and outline for how to place nuance, relationality, and contextualization before and above the systematization of rules or tests, in determining or applying labels of consciousness.

Keywords: Consciousness, Machine Consciousness, Philosophy of Mind, Phenomenology, Bodyminds

1 Introduction

If we would claim to consider the notion of nonhuman consciousness, we must necessarily do more than engage questions such as "under what rubric and by what metric would such a thing be possible?" Any machine consciousness we manage to generate will be simultaneously like and unlike the humans who form the basis of its generation. It will be like humans in that it will be made from and by humans, and will thus have their perspectives and biases; unlike in that it will not be made of the same constituent components as humans, and will not intersect with and relate to the world in the same way. To place human-like expectations on such a consciousness would be to fundamentally disrespect the alterity of that consciousness—to have decreed it as "other," and declared that otherness as unacceptable.

To truly be said to have sincerely considered the concept, we must examine the sociological, moral, political, legal, and relational implications of a nonhuman, nonbiological generated consciousness, and to consider those implications, we must look at what historical, often deadly precedents we have for how we have responded to being confronted with unexpected forms of consciousness, both nonhuman and human. In this paper I argue that many western conceptions of consciousness, though diverse and inclusive of different ***clusters*** of entities, are all exclusionary of at least one category which, in humans, we would say are conscious and worthy of consideration. To that end, I propose examining both nonwestern systems and disciplines such as race gender and disability studies for what new perspectives on consciousness they may grant.

Further, as consciousness is rather nebulously connected to the legal notion of personhood, and since personhood is a primary category by which we currently grant rights, consideration, and political power, we must fundamentally reckon with both consciousness ***and*** personhood, and how power, consideration, protection, and breathing space flows to those entities (actual and potential), under their remit. I will examine the implications of the legal notion of personhood and its flaws, arguing for a revision if not complete overhaul. Additionally, I provide support for the notion that consciousness is not any one exclusive thing, and that while theories of consciousness provide rubrics by which to say what may or may not be conscious, they still exclude types of existence embodied by individuals whom we ***would*** say are conscious.

This paper looks not only to philosophy, computer science, and neurobiology, but to phenomenology of race, gender, and disability, and to the historical contexts in which humans and nonhuman animals have been granted personhood status. I also explore questions of intersubjectivity, and the shared creation of knowledge, understanding, and reality itself. Overall, I argue that it is only through creating an intersubjective account of knowledge and consciousness that we can account for the many kinds of lived experience present in humans and nonhuman animals, and further guide our paths in the creation of precedent by which to prepare ourselves to encounter nonhuman, nonbiological minds. Ultimately, this work proposes an interdisciplinary program for seeking, categorizing, and engaging multiple kinds of minds, whereby differing insights might be leveraged against each other to provide room to synthesize new perspectives. If we ever hope to relate to and engage conscious machines, we must compile toolsets which will help explore these minds' similarity to ***and*** their alterity ***from*** the minds that make them.

2 On Consciousness, "Artificial" and Otherwise

Before we can discuss the potential consciousness of machines, we have to ask, what is consciousness? And in asking that, we must be aware that the definition of consciousness we use will necessarily change what we will or even ***can*** identify as a conscious agent, at the outset. There are multiple differing philosophical and neurobiological frameworks for consciousness, such as mind-body dualism, nonconsciousness, functional consciousness, and phenomenal consciousness, embodiment, and extended mind, and each has different implications and outcomes when used as the rubric by which we search for or set out to classify conscious entities. Even more fundamentally than this, however, we must reckon with the implications of using the term "Artificial Intelligence" in seeking to discuss potentially conscious machines.

The word “artificial” prejudices the discourse about the status of these minds by placing them as “unreal” and “less than.” At the outset, this framing places any created or generated intelligence on the defensive, forcing it to support its own value and even the very reality of its existence and experience [1].

Though they may certainly have been intentionally formed, and with an eye toward their potential capabilities, there exists no measure to reliably argue the “artificiality” of an entity’s consciousness. Tests and metrics for consciousness and mindedness have problems ranging from registering dead fish as conscious to not registering living, aware humans as such. For this reason, I prefer the terminology of “Autonomous Created Intelligence,” first put forward by Jamais Cascio [2], “Autonomous Generated Intelligence” (AGI), as coined by researcher Emily Dare, in conversation, or simply “machine minds.” Additionally, as we ask if there is there any such thing as “what it’s like” to be a mind, or whether we will ever solve the problem of other minds, or whether we should even be trying to do so, we must question the fundamental assumptions on which our ascription of consciousness are and have been based.

The concept of “phenomenal” consciousness holds that humans are more than just symbol-crunching machines. That human experience is the experience ***of*** something, and that there is something it “is like” to be a mind, and to experience things. This is often called “qualia,” deriving from “subjective” or “qualitative” experience. One of the most famous explorations of this idea is Thomas Nagel’s seminal paper “What Is It Like To Be A Bat?“ In this, he argues that a human being, even a celebrated bat biologist, will never truly understand what it is like to have the lived experience of a bat, because humans do not exist with the set of physiobiological constraints that bats do. Bats live in low light and hang upside-down in groups to nest, eat fruit, drink blood, or hunt via echolocation; and no human does all of them at the same size, scale, and combination that a bat does.

In this context of this work, phenomenology is taken to mean lived experience, the felt-sense experiential knowledge of the world, from within particular contexts such as race, gender, disability, age, sexuality, etc., any or all of which can be modulated by external and internal factors. This draws from work done by theorists of race, gender, and disability to build upon the foundation laid by philosophers like Edmund Husserl, in his *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy*, and Maurice Merleau-Ponty in his *Phenomenology of Perception* [3]. Likewise, Intersubjectivity pertains to knowledge and experience shared between individuals and groups of individuals who regard each other as legitimate subjects, rather than as objects, creating a shared reality from which to act.

For this purpose I believe we must investigate nonwestern philosophies with radically different conceptions of epistemology and consciousness than are often found in Western schools of thought. While the nonwestern turn is not without its faults (the Buddha didn’t believe women could attain enlightenment, for instance), these systems provide a foundation from which to explore and they serve as object lessons for the overarching thrust of this project: That different systems of knowledge will provide different internally consistent answers in different situations, and different phenomenological experiences will produce both different pictures of the world and different systems by which to navigate them.

In order to fully understand what it means for different lived experiences to produce different kinds of consciousness, we have to understand that culture and society form an extended mind which forms and shapes our physical forms and consciousness into what disability studies scholar Margaret Price terms "bodyminds" [4], and provides us with the template via which to act in the world. In contemplating the limits of that action, the limits of internal identity, and the boundaries of self and other, we may look to examples such as the Ship of Theseus problems and questions of proprioception, time, flow, and nownesss, all of which seem to teach important lessons about the unity of mind and body. Here, we may explore how the cases of Ian Waterman [5], conjoined twins Abigail Loraine "Abby" Hensel and Brittany Lee Hensel [6], and others with "nonstandard" configurations of physiological intersection with the world, call our simplistic, clear-cut notions of self into question. Through these cases, we can see that any sets of physiological and neurological correlates of consciousness may have circumstances in which they do not adequately describe what we observe.

The particular arrangements that we tend to call human consciousness are, by definition, *correlational* and so that means that there may not be any particular thing that makes humans a special case of consciousness. This correlational nature also seems to indicate that there is no single, particular organizational structure that is universally necessary for all kinds and instantiations of consciousness. We can see evidence for this in the history of studying nonhuman animal minds, such as in the work of Peter Godfrey-Smith [7], spider cognition [8], corvid cognition [9], and more. But we must also consider the many times in *human* history where "nonstandard" minds have been categorized as less than "fully human"—women have been institutionalized for being outspoken; Black Americans enslaved, sterilized, and experimented upon against their will or without their knowledge [10]; autistic folx, people with Down Syndrome, and other neurodiverse populations have been said to "not really feel pain" or subject to vast eugenics campaigns [11]. In each of these cases, reportage from various sets of lived subjective experiences was discounted in favor of a perceived "right kind" of way to be, and that discounting resulted in the systematic degradation of entire categories of people.

No consensus, nor even an intentionally divergent multiplicity, exists regarding the question "What Are We Trying to Build?" Do we want better tools? Or do we want to build new minds? Some want the former, others want the latter, and there is no communication as to which project should receive precedence, and why. And while this distinction might not matter in the consideration of any other technology, if we want, or accidentally ***happen*** to create conscious machine minds, we most definitely ought not to treat them ***as mere tools***. Human history teaches us that a mind which recognizes itself as being treated as a tool, without regard for its sense of itself as an agent or a subject, will likely rebel—and is that mind not right to do so? In historical uprisings of those humans who have been continually oppressed, tortured, degraded, killed, experimented on, or enslaved, we more often than not hail them as heroes for demanding their rights to exist. To this end, we must do the work to be clear about our aims, well in advance.

In order to fully consider how a conscious machine might exist in the world, we need to think about a plurality of types of consciousness***es***, knowledg***es***, and intelligenc***es***, rather than deploying “consciousness,” “knowledge,” or “intelligence,” as though they were singular concepts. In this project, however, we run into the danger of anthropocentrism and its correlate, what Ashley Shew calls “The Human Clause” [12]. Considerations of consciousness lead, again, to notions of personhood, which are often assumed to be directly equivalent to some estimation of humannesss. This, however, is just another variation of the same “Right Kind” Bias, mentioned above: the tendency toward the belief that there is a “Right Kind”…of mind, …of body, …of skin, …of gender, …of sexuality, …of thought, …of life, …of religion. And the belief that these “right kinds” of traits exist, at all, tends to give rise to a belief that only those who embody those kinds are “really human.” Further as humanness is the basis for analogy on which we tend to judge personhood, then anything that isn’t “the right kind” of human then necessarily strains its claim on the title “Person,” and is thus potentially unworthy of being seen as a conscious mind [13]. We must come to understand that there is no single “right kind” of consciousness—which means that there cannot be any single ***test*** for consciousness, either.

3 Personhoods

In the case of potential machine minds, tests such as the Turing or mirror tests should not only be considered woefully inadequate to the task of identifying minded machines, but it should be understood that this test misses and will continue to miss some humans who are not considered “normal” [14]. As mentioned above, there are a statistically significant number of humans who fail tests devised to assess “normal” human consciousness, because they were prejudicially excluded from the definitions of “consciousness,” “humanness,” and “personhood.” Western definitions of personhood have included both the social (de facto) definition, wherein what counts as a person is who- or whatever is accepted by the local or wider community, and a legal (de jure) definition, wherein a person is who- or whatever has the protections of personhood under the law.

But de jure protections without de facto acceptance will still result in ostensible persons being treated as non-persons. Even philosophical definitions of “persons” have, at least, tended to depend on hierarchical rankings of entities [13]. African Americans, women, the disabled, the neurodivergent, and LGBTQIA phenomenologies have all been deemed illegitimate candidates for personhood at some point, and some still are, to this day. These groups received or still receive no legal protections and it was not until we fought to change the framework of our measurements that that we ostensibly obtained said protections. I say ostensibly because, again, some in those groups still get killed by states apparatus with no substantive repercussions.

Our legal measures for assessing persons often fail in the face of the multiform externalities of what it means to ***be*** a person. Different societies around the world categorize different species, objects, systems, and groups as different types of persons, including refusing to accept the reality of the lived experiences of groups and types of humans, regardless of bodies of scientific evidence that say we should. No single test or standard could be dispositive of every type of consciousness or mind which might need inclusion under the umbrella of "personhood." But there are nonwestern definitions of personhood, as well. The rivers of the Maori people are understood as Natural and Spiritual Ancestors [15]; Shinto Kami as Concretions of Natural/Spiritual Energy; the Jains understand all Life to be worthy of care, dignity, and respect. Yet even with these longstanding examples, Western-style legal systems hold a requirement that individuals or groups meet ***their*** standards before widespread acceptance of a new candidate for personhood can even be considered.

4 Learning From Each Other

There are many very good reasons to think that the human perspective for understanding consciousness, sentience, or even cognition is woefully inadequate to the task of thinking about even whether other humans and nonhuman animals have these things, let alone whether non-animal or even non-biological entities might. Again, the Maori, Hindus and others think rivers and the whole natural world can act, suffer, and have sentience, and thus are conscious [15, 16]. To adherents of these beliefs and members of these cultures, most Western measures for assessing all of these traits are woefully incomplete and anthropocentric. And what if they're right? What can we do to think about this differently and what changes in our assessments, when we do?

We must work to believe in the existence and reportage of those different from us, and to do so, we might start by examining the generation of identity, belief, and knowledge, with reference, again, to the phenomenology of, race, gender, sexuality, disability, and age. We can use these factors to build a basis for intersubjective phenomenal knowledge, recognizing and accepting the shared contexts and corroborations of internally consistent understandings between those accepted as persons. The goal should be to not treat consciousness as a zero-sum game, one without the possibility of multiple correct perspectives or multiple right answers. If, instead, we say that many though not all perspectives can be right, then we can learn much more—we can have many new knowledges. With these knowledges, we can explore the crucial difference between the statements "there can be more than one right kind of answer," and "all answers are equally right." Because while all descriptions of consciousness and knowledge come from the same place (i.e., we make them up to explain the world we experience), some descriptive systems make it easier for us to oppress and murder each other, while other systems seek to help us all to flourish.

Many, including some in the present day transhumanist movement, actively support eugenics-like initiatives to make sure that "only the best" can reproduce or survive. While wanting "the best" seems an innocuous enough idea, those who hold it have often ended up calling for the destruction of entire categories of people, or at the very least believing some people to be so inferior that either their destruction or the limitation of their agency poses no moral hazard [17]. Philosophers and scientists who continue to think there must be one and only one correct way for consciousness to exist in the world open the door for anyone who can string a hateful syllogism together to say their views are supported by "the best minds."

Instead, we must start from the position of believing that there are other kinds of minds, even among humans, and hence there can be no singular default type, no normative ideal. Here, I propose an interdisciplinary system combining the lenses of feminist epistemology, standpoint theory, and the other theoretical frameworks mentioned above, as these perspectives preference the perspective of what Donna Haraway terms “The Subaltern,” those people who are most often ignored or unlikely to be perceived as knowledge-holders [18]. Though I preference the fields of race and gender studies, disability studies, philosophy of mind, philosophy of technology, and animal studies, this is not the only path to interdisciplinarity. By combining the methodological tools of scientific, technological, and the humanities disciplines, we give ourselves a better opportunity to understand those lives and consciousnesses which contravene our assumptions of what qualifies as conscious, or even “normal.” To do otherwise is to remain in real, recurrent danger of both doing harm to different humans because they “aren’t really people,” and failing to recognize nonhuman consciousnesses who experience their lives in ways we classify as “illegitimate” and who thus possibly suffer in ways we don’t count as “really suffering.” If we do not accept and seek to understand the lived, phenomenological knowledge of their experience, how would we ever know?

5 Conclusion

For decades—if not centuries—humans have contemplated how to make conscious machines, without a full recognition of what consciousness is, or an honest appreciation of the precedents for how we have tended to politically, legally, morally, and sociologically relate to new candidates for consciousness. If we would sincerely consider what it would take to create a machine mind, then we must also consider what it would mean for us to ***accidentally generate*** one. Any AGI will both resemble and be distinct from its human creators, and we must work to respect their fundamental alterity. To do this we must work to understand different human and nonhuman bodyminds, paying special attention to those lived experiences which have most often been oppressed, disregarded, and marginalized, as those phenomenologies will have developed unique epistemological strategies to understand and navigate the world.

Subaltern phenomenologies generate knowledges and perspectives from and about which we can learn to more quickly, agilely, and robustly adapt our notions about how to create and—more importantly—***recognize and understand*** new kinds of minds. This increased agility, speed, and robustness of thought will, in turn, prepare us to accept and respect the reportage of potentially conscious machines, reducing the likelihood of our causing them to suffer. More succinctly: If we ever want to create and know a conscious machine mind, we should first listen to those living people who are different from us and who have been systemically prevented from speaking to us, because they will know a lot that we don’t about consciousness, social and legal personhood, and being made to both submit to testing about and argue the validity ***of*** their own lived experiences. And both they and future minds would likely very much appreciate it if we heeded them.

References

1. Williams, Damien Patrick. “Strange Things Happen at the One Two Point: The Implications of Autonomous Created Intelligence in Speculative Fiction Media;” The Machine Question Symposium. July 2012.
2. Cascio, Jamais. “Cascio’s Laws of Robotics.” Bay Area AI MeetUp. Menlo Park, Menlo Park, CA. 22 March 2009. Conference Presentation.
3. Browne, Simone. *Dark Matters*. Durham, NC: Duke University Press, 2015; Husserl, Edmund. *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy*. Hague: M. Nijhoff, 1982; Merleau-Ponty, Maurice. *Phenomenology of Perception*. London: Routledge, 1962.
4. Price, Margaret. “The Bodymind Problem and the Possibilities of Pain.” *Hypatia* 30. 2015.
5. Cole, Jonathan and Waterman, Ian. *Pride and a Daily Marathon*. MA: MIT Press. 1995
6. Schrobsdorff, Susanna. “Reality’s Believe It or Not”. Newsweek.com. February 23, 2008. http://www.newsweek.com/realitys-believe-it-or-not-93725.
7. Godfrey-Smith, Peter. *Other Minds: The Octopus, The Sea, and the Deep Origins of Consciousness*. New York, Farrar, Straus and Giroux, 2016
8. Japyassú, H.F. & Laland, K.N. “Extended spider cognition.” *Animal Cognition*. (2017) 20: 375. https://doi.org/10.1007/s10071-017-1069-7.
9. St Clair, James J H and Christian Rutz. “New Caledonian crows attend to multiple functional properties of complex tools” *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* vol. 368,1630 20120415. doi:10.1098/rstb.2012.0415
10. Tuskegee University, “About the USPHS Syphilis Study.” https://www.tuskegee.edu/about-us/centers-of-excellence/bioethics-center/about-the-usphs-syphilis-study.
11. Ustaszewski, Anya. (2009). “I don’t want to be ‘cured’ of autism, thanks.” The Guardian. https://www.theguardian.com/commentisfree/2009/jan/14/autism-health; Verbeek, Peter-Paul. “Obstetric Ultrasound and the Technological Mediation of Morality: A Postphenomenological Analysis.” *Human Studies*, Vol. 31, No. 1, *Postphenomenology Research* (Mar., 2008), pp. 11-26; Springer. http://www.jstor.org/stable/40270638; Kafer, Allison. *Feminist, Queer, Crip*. Bloomington: Indiana University Press, 2013
12. Shew, Ashley. *Animal Constructions and Technological Knowledge*. Lanham, MD: Lexington Books, 2017.
13. Cf. Mary Midgley, René Descartes, John Locke, Harry G. Frankfurt, and many others.
14. Anderson, J.R. & Gallup, G.G. “Mirror self-recognition: a review and critique of attempts to promote and engineer self-recognition in primates.” *Primates* (2015) 56: 317. https://doi.org/10.1007/s10329-015-0488-9
15. Roy, Eleanor Ainge. “New Zealand river granted same legal rights as human being.” The Guardian. March 16, 2017. https://www.theguardian.com/world/2017/mar/16/new-zealand-river-granted-same-legal-rights-as-human-being
16. “[Court Declares] India’s Ganges and Yamuna rivers are ‘not living entities.’” BBC News. July 7, 2017. https://www.bbc.com/news/world-asia-india-40537701
17. Singer, Peter. *Practical Ethics*. New York: Cambridge University Press. 2011; Singer, Peter. “Taking Humanism Beyond Speciesism.” *Free Inquiry*, 24, no. 6 (Oct/Nov 2004), pp. 19-21.
18. Haraway, Donna. *Simians, Cyborgs And Women: The Reinvention Of Nature*. NY: Routledge. 1991.