

The Oracle

SUMMER 2026
ISSUE 19



PHILOSOPHIA
YORK UNIVERSITY

JACK MARSHALL
UNTITLED

The Oracle

York University's Undergraduate Philosophy Review Journal
Issue 19 (Summer 2026)

***The Oracle* Editorial Board**

Mirabelle Martin
Thomas Michael Covert-Pareja
Adam Jackson
Lucas Firas Kayssi

Front Cover Artwork

Jack Marshall

Philosophia Executive Team

Syeda Gul Zahra
Harrish Santhuruselvan
Janan Kandeepan
Harlin Virk
Parmeet Singh
Sana Paracha
Grace Williams
Alisha Moin
Medya Ekhlaspour
Romi Mikhailov
Naina Puri
Calleigh Wrobel
Nitzana Philip
Dominique Lindo
Mirabelle Martin
Thomas Michael Covert-Pareja

Table of Contents

Acknowledgments 5

Letter from the Editors 6

Thomas Michael Covert-Pareja, Mirabelle Martin

ESSAYS

In Perpetual Motion: A Dialectical Reading of Michael Della Rocca 8

Kore Moros Scheibe, York University

Edited by Adam Jackson

Dignity to Forests: Introducing the Notion of Honor towards Old-Growth
Forests in the “War in the Woods” 25

Kyara Lalli, University of British Columbia

Edited by Thomas Michael Covert-Pareja

VTubers: Para-Social Relationships with Simulacra on Virtual Hyperreality
50

Andrew Elvio Febrian Tjioe, York University

Edited by Thomas Michael Covert-Pareja

From Reasons to Reason-Shaped Prose 65

Mohiz Muhammad, University of Windsor

Edited by Lucas Firas Kayssi

The Origins and End of God: An Essay Exploring Man’s Power over God
90

Parsa Heidari, University of Toronto

Edited by Mirabelle Martin

**Ignorance is Risk: The Ethics of Developing Information Technologies as
Forms of Life 101**

Audrey Kruger, University of British Columbia

Edited by Adam Jackson

**I Am, But Do I Think? On the Historical and Ongoing Devaluation of
Philosophers as a Symptom of Thoughtlessness 122**

Paloma Claire Figueroa, McGill University

Edited by Thomas Michael Covert-Pareja

Marx and Engels on Nature: Through Darwin, against Malthus 173

David de Graaf, University of Ottawa

Edited by Mirabelle Martin

Essentialist (Mis)givings 196

Yǔhuán “Albert” Xiè, University of Toronto

Edited by Mirabelle Martin

WRITTEN CREATIVE WORKS

David King 219

Camron Lee Welsman, York University

The Monster 224

Mister Glass

The Philosophy of Environmental Design 226

Atakan Elalmis, York University

PHOTOGRAPHY AND VISUAL ARTWORK**Rebirth of the Forest 228**

Tatijanna Martel, York University

The Questions Inside of Us 230

Calleigh Susan Elizabeth Wrobel, York University

The Witness 232

Naina Puri, York University

Untitled 234

Jack Marshall, University of British Columbia

Acknowledgments

Look what they got me writing...

This edition of *The Oracle* was made possible mainly through the time, efforts, and commitment of our 2026 Editorial Board: Mirabelle, Thomas, Adam, and Firas.

Mirabelle, our whimsical linguistic-cel, thank you for asking critical questions and for prioritizing the quality of this journal. Tom, our night owl with 12 hours of sleep debt, thank you for always checking in and keeping the team on track. You two make it hard to believe you haven't done this before. You're the life of *The Oracle*!

Adam and Firas, thank you for joining (and saving) the board. *The Oracle* has benefited especially from your collaboration as PhD students. We're very lucky to have you as friends and supporters, not just of the journal, but of Philosophia too.

The York University Undergraduate Philosophy Conference is, as always, backed and supported by the administrative staff in our Department of Philosophy, and Professor Giudice and Professor Waring, whose enthusiastic support and endorsement has kept us going strong since the conference's resumption.

We're grateful for each and every student that submitted a piece to this edition of *The Oracle*, and especially for our presenters: Kore, Kyara, Andrew, Mohiz, Parsa, Audrey, and Paloma - we hope this experience was fun and challenging in all the good ways.

This I say all the time, but I hit the jackpot with my team this year. Aside from carrying Philosophia, you (yes... all 15 of you) are easily the most driven, capable, and fun group of people I have gotten to know in my university years. I'm fortunate to be working with you and can't wait to make more great memories together.

Finally, for giving so much to this journal for four long and turbulent years, a huge shout-out goes to Micheal Habib, *The Oracle's* former Editor-in-Chief and my good friend. I hope you feel fulfilled seeing all this passion for and dedication to *The Oracle*.

I sure do!

Syeda Gul Zahra
President, Philosophia



Letter from the Editors

Upon announcing this year's *The Oracle* theme, we were met with many questions—*how broad is the theme? How strict is the theme? What does the theme mean?*

On **beginnings and endings**, we present a vast collection of works: from the study of where the self ends and the public persona begins with virtual performers to the examination of how nature becomes a subject of honor when observed by society, and how the beginning of a new age of AI and devaluation of wisdom may instead be a phenomenon occurring since the conception of the practice of Philosophy.

This edition is an opportunity to look at the thoughts of a fresh generation of philosophers, and it is no surprise that many of them are occupied with the ethics and metaphysics of virtual realities, Artificial Intelligence, and the environment.

While some of this year's submissions explicitly explored origins and ends, others engaged the theme in more oblique ways: through allusions to nostalgia; through the defense of that which seems lost; by revealing an escape-hatch—a door, a threshold—all the while questioning what it would mean to stand before it, step through it, or linger between. Clarity and consensus are no guarantee. We must acknowledge that our discussions occur in context—shaped by our present, our past, and our futures: beginnings, ends, and perhaps always, necessarily, both.

In a year shaped by accelerating global change and instability, this issue of *The Oracle* showcases the resilience of critical thought and discussion. We hope our readers take from it what they need and continue the work of reflection beyond these pages.

With a resounding “thank you” to the Philosophia team, the rest of *The Oracle* team and, of course, our brilliant writers and creatives, we invite you to read *The Oracle's* 2026 edition. Because what good is art without an audience, or science without peer review?

Thomas Michael Covert-Pareja
Editor-in-Chief, *The Oracle*



Mirabelle Martin
Editor-in-Chief, *The Oracle*



The Oracle

In Perpetual Motion: A Dialectical Reading of Michael Della Rocca

Dialectics and monism seem at first glance like contradictory concepts: di-, a pair, both, double; mon-, one, single, sole. In this reading of Michael Della Rocca's book-length argue for monist philosophy, The Parmenidean Ascent, I argue that in trying to achieve monism from a starting point of relational metaphysics, we are inherently engaging in a dialectical exercise, and that such exercise is in fact more valuable than either our starting point or end goal. To do so, I describe Michael Della Rocca's argument for radical monism, putting it into dialogue with Hegel's dialectics. In doing so, I argue that the process of arguing for radical monism is more key to The Parmenidean Ascent's insight than monism by itself. Not only that, but it is our only option, as I argue that Della Rocca's Ascent renders both its starting and end points invalid, and leaves only the motion between them. I then demonstrate how we might apply such an understanding of perpetual argument by connecting it to the method of phenomenology, specifically as it is formulated in the work of Maurice Merleau-Ponty. Finally, I conclude that this perpetual incompleteness is a feature, not a bug: in valuing a process of constant ascent, we can find value in the fact that philosophical inquiry can always be taken further, that there is always more to be found and to know. To support that such a state is more insightful than completion, I highlight examples of insights derived from similar perpetual dialectic motion in contexts other than metaphysics.

Dialectics and monism seem at first glance like contradictory concepts: di-, a pair, both, double; mon-, one, single, sole. In this reading of Michael Della Rocca's *The Parmenidean Ascent*, I argue that in trying to achieve monism from a starting point of relational metaphysics, we are inherently engaging in a dialectical exercise, and that such exercise is in fact more valuable than either our starting point or end goal.

To do so, I first briefly describe Michael Della Rocca's argument for radical monism, and put it into dialogue with Hegel's dialectics. Under this framing, I show that the process of arguing *for* radical monism is more key to *The Parmenidean Ascent's* insight than monism by itself. I support this reading by asserting that Della Rocca's *Ascent* succeeds in rendering relational metaphysics invalid, but does not make a standalone case for the validity of monism, only for the motion between the former and the latter. I then demonstrate how we might apply such an understanding of perpetual argument by connecting it to the method of phenomenology, specifically as it is formulated in the work of Maurice Merleau-Ponty.

Here, I argue that just as Merleau-Ponty's phenomenology requires the continuous existence of a differentiated world which must be rendered undifferentiated in order to perceive it, Della Rocca's *Parmenidean Ascent* requires the pre-existence of relational metaphysics, as it is only through relational logic that Della Rocca's monism is achieved. I show that this, when combined with Della Rocca's sound argument for the invalidity

of relational metaphysics, forces us to simultaneously move towards monism and acknowledge that we can never *reach it*, making the *Ascent* a perpetually incomplete, but worthwhile process.

Finally, I conclude that this perpetual incompleteness is a feature, not a bug: in valuing a process of constant ascent, we can find value in the fact that philosophical inquiry can *always* be taken further, that there is always more to be found and to know. To support that such a state is more insightful than completion, I highlight examples of insights derived from similar perpetual dialectic motion in contexts other than metaphysics.

Michael Della Rocca's Monist Argument

Though Michael Della Rocca's Parmenidean Ascent explores several different philosophical concepts, ranging from knowledge and substance to action and philosophy itself, he takes roughly the same approach in engaging with each of them. First, he lays out the standard rationalist and relational metaphysics understanding of the concept, as posited by a variety of philosophers (such as Locke, Leibniz and Descartes, in the case of substance)¹, and shows how all such views ultimately depend on relations between differentiated subjects. Though this

¹ His engagement is broader (and more modern!) when it comes to other sections, such as on the concept of action or knowledge.

is not usually seen as a problem, Della Rocca goes on to argue that there is no way to define relations that does not violate the Principle of Sufficient Reason, or PSR, which rationalist metaphysics claims to accept. To actually accept the PSR, Della Rocca argues, we must perform an *ascent*, in which we discard our prior notions of differentiation and relations and accept a radical monism towards metaphysics (Della Rocca, 2020, pp. 77–79). To break down each of these steps, let’s examine how Della Rocca tackles the idea of *substance*.

Our first step is laying out the mainstream view of substance in analytic metaphysics: (i) some things are *substances*, (ii) a *substance* is a different thing than *action* or *knowledge*. We will also usually want there to be *different substances*, not just one. (Della Rocca 2020, 53–56) Finally, most² philosophy works under the Principle of Sufficient Reason: we cannot simply posit things without explaining *why* they are so, e.g., we need a *reason* why something is a substance instead of an action. (Della Rocca, 2020, xiv (proem)). Locke, for example, asserts that substances are a union of qualities (e.g., a stone unifies size, mass, color, etc.), and that there must be some reason that this is the case, though he self-admittedly isn’t sure what it is (Della Rocca, 2020, p. 53).

² Though not all philosophy does or must, as Della Rocca (2020, p. 230) notes when discussing Hume and other “PSR-deniers”.

The dependence on relations lies in two separate places here. The first instance is when we say that there are *different* substances: for example, if we say that silver *is a different substance* than gold, then we are defining silver *in relation* to gold, and silver is thus not a simple, self-justifying substance (Della Rocca, 2020, p. 11). Even if we dispute this³, there is a much more fundamental dependence on relations in this view. The PSR demands we not only say that silver is a substance, but that it is so *because of* something. This is a fundamentally relational statement: we are saying that we can only define silver through its relation with *something else*, be it material composition, a broader set of its qualities, etc. (Della Rocca, 2020, pp. 56–57).

Under the PSR, however, we cannot simply state that things *are* without justification, and this includes relations. If we say that physical objects are substances because they are composed of some arrangement of matter, we must explain just how it is that arrangement of matter makes something a substance. We of course will then need to explain how an arrangement comes to be in the first place, and so on (Della Rocca, 2020, p. 56).

Here we come to Della Rocca's (2020, pp. 60–62) second step: if we demand that all things must have a reason that they are so, we must continuously also explain *why* those

³ For example, we might say that the difference between gold and silver exists *after* their respective definitions, not as a part of them (Della Rocca, 2020, p. 11).

reasons are tied to the things they explain, in an infinite regress. Thus: to say that **A** is explained by **B** is to say that there is some explanatory relation **R** between **A** and **B** ($R \rightarrow B$ explains A). But **R** itself requires another explanatory relation, **R**₁ ($R_1 \rightarrow R \rightarrow B$ explains A), which itself requires an explanation, ad infinitum (Della Rocca, 2020, p. 64). We could short-circuit this process by saying that **R** is justified by the fact that **A** and **B** are related, but this is equivalent to saying that **R** exists because **R** exists: $R \rightarrow B$ explains $A \rightarrow R$. This does not seem particularly illuminating or in the spirit of the PSR (Della Rocca, 2020, p. 67).

If we are convinced of this second step, we are left with two options: abandon the PSR, or abandon relations, and with them the power to differentiate between things. To choose the latter is to accept what Della Rocca calls *radical monism*, or *RM*: the view that we cannot make any rational case for differentiation between concepts or instances of concepts⁴. We cannot rationally justify why a particular thing is a substance, or that there are distinct substances, and therefore, we cannot say that there are particular substances at all. We might⁵, however,

⁴ Though the option of abandoning the PSR is out of scope for this essay, Della Rocca does discuss whether we are able to do so (and what it might cost us) on p. 258.

⁵ Della Rocca (2020, pp. 11–12) asserts that we can only say that *all is substance* instead of *all is not substance*, as the latter is a relational statement, like saying that silver *is not* gold, while the former isn't. I am not fully convinced by this precedence of being over non-being, and revisit it through my introduction of Hegelian dialectics.

be able to say that *all is substance*, though we can equally say that *all is action*, as our acceptance of RM makes us unable to distinguish substance and action in the first place. This acceptance is what Della Rocca terms the *Parmenidean Ascent*: to leave all previous notions, including ones we used to argue for RM, behind, and embrace the monist perspective (Della Rocca, 2020, p. 219).

Can we be truly monist | Can we truly not be monist?

One part of that ascent, though, might seem immediately troubling: Della Rocca's ascent includes throwing away the very tools that justify making an ascent in the first place, the sufficient reason to embrace monism. By his own admission, his arguments for RM are only possible with the starting point of relational metaphysics, and in undermining that starting point, he undermines his own path out of it, too. That, however, doesn't seem to worry him: why would we care about our path to RM, considering we've already arrived at it? Here, Della Rocca (2020, pp. 221–223) compares the Ascent to the (in)famous end of the *Tractatus Logico-Philosophicus*, where Wittgenstein asserts that the conclusion of his work makes all of his previous claims nonsense, but that it doesn't matter: we've already used them to achieve a new paradigm⁶, and have

⁶ Here, I am importing Della Rocca's (2020, p. 221) own statements on Wittgenstein, and my discussion of it is only applicable within the context of Della Rocca's

no further need of them. Relational metaphysics is a *ladder* with which we *ascend* to monism, after which we are done with it, free to discard it or scrap it for parts.

But precious little comes *after* the ascent in Della Rocca's work. There are tentative moves towards a "non-relational seeing", loose assertions that we are now free to use the language of relational metaphysics as arbitrarily as we want, without the (to Della Rocca, impossible) burden of making it coherent under the PSR. Perhaps it is unfair to ask for a clear view of this new paradigm in the immediate aftermath of reaching it. However, I see it as a clear sign that we are *not done climbing*. In fact, in the last paragraphs of Chapter 9, Della Rocca acknowledges that we might only be able to accept RM partially and gradually. But he still asserts that it is a standard we should aspire to (Della Rocca, 2020, p. 225).

This one-paragraph acknowledgement, to me, differentiates (sorry!) Della Rocca's *Ascent* from Wittgenstein's ladder: the latter assumes that we can reach an endpoint wherein we are safe to discard it⁷, whereas the former seems unable to ever be fully complete, either practically or because in reaching

own work. Readings of the *Tractatus* are also a heavily contested subject, with Della Rocca's view being self-admittedly aligned with the Resolute reading of Cora Diamond (1991) and James Conant (1992).

⁷ As previously, this is based on Della Rocca's own stated views on Wittgenstein, and is only applicable within that limited scope.

such a state we would be losing our ability to assert that we have moved at all.

Hegelian Dialectics & Della Rocca

What are Hegelian Dialectics?

To understand what this inconclusiveness might mean for Della Rocca's ascent, it is helpful to put it in dialogue with another notoriously radical view of metaphysics: Hegel's dialectics.

Hegel first posits that *pure being* and *pure nothing* are both states of undifferentiation and indeterminateness. As with Della Rocca's radical monism, we can make no assertions about their specific contents, only that *being is*, or *all is being*. But the same could be said of nothing, itself another undifferentiated state. Therefore, *all is*, and at the same time, *all isn't* (Hegel, 2014, chap. 1, sec. A-B).

To Hegel, determining in favour of either would transform them into determinate and differentiated concepts, which would themselves require the kind of grounding that seems unfeasible (as we have seen in Della Rocca's arguments for RM). To not do so, however, is impossible, as both concepts assert that they are distinct *from each other*: either *all is* **or** *all isn't* (Hegel, 2014, chap. 1, sec C.1-4).

Instead of fighting this contradiction, Hegel argues we should accept that we can only know that the contradiction is constantly resolving itself: *being* must turn into *nothing*, and *nothing* must become *being*. We cannot understand either end-state, and regardless, it would transition into the other as soon

as it started; there is nothing but the constant movement between them, or as Hegel calls it, *becoming* (Hegel, 2014, chap. 1, sec 2).

The Dialectics of Ladders

Applying this mode of thought to Della Rocca means asserting that neither radical monism *nor* relational metaphysics are comprehensible and/or valid positions by themselves, i.e., that both distinction and indistinction are nonsense. In doing so, I assume acceptance of Della Rocca's claims about the invalidity of relational metaphysics *and* the self-undermining nature of his own monist account.

There is one additional assumption, however. In Hegel's framing of dialectics, the motion between nothing and being is driven by the fact that both propositions are defined by (i) their indeterminateness and undifferentiation within themselves but also by (ii) their complete opposition to *each other*, which yields a constant state of contradiction. This is in contrast with Della Rocca's (2020, pp. 11–12) assertion that nothing is defined as a negation of being, while *being* itself is conceptually self-sufficient, i.e, does not require a negation in its definition.

This is not justifiable except by a pre-existing preference for *being*. If we do not presume the existence of *being*, we can easily define *nothing* as requiring no negation of it. Therefore, we can either assume only one exists (in which case, we cannot justify picking *being* over *nothing*), or assume that both exist and are defined as negations of each other (as in Hegel's framing). A similar objection to Della Rocca's preference for being has been raised by Filippo Casati (2025, sec. 6).

If we accept, then, that (i) Della Rocca's claims about the invalidity of relational metaphysics hold, (ii) but are also self-undermining, and that (iii) we may not give precedence to either relational metaphysics or monism, we are left with nothing but the process of moving between the two states, of performing a Parmenidean Ascent (and also a Descent). Interestingly, similar readings have been made of the aforementioned Wittgensteinian ladder: instead of understanding Wittgenstein's plea as one that asks us to discard all his previous arguments, we might see it as asking us to constantly engage with them, to never stand still and accept them but be in a constant state of "intellectual restlessness" (as described by Ben Ware (2015, pp. 49-51).

With this framing, I now aim to establish (a) why ascending is necessary, (b) why it is a process that can never truly be complete, and finally (c) why this state of constant ascent is a state we might find comforting instead of frustrating.

Why Ascend?

A need to erase differences between ourselves and the world is not only a core feature of Della Rocca's work, but also that of the French philosopher Maurice Merleau-Ponty. In *Phenomenology of Perception*, Merleau-Ponty (2012, pp. 28-30, 69) argues that the objects we perceive necessarily pre-exist our seeing of them: objects are differentiated and external to us. Equally core to Merleau-Ponty's argument, however, is that in perceiving an object, we seek to erase that differentiation; in fact, this seeking itself is what perception is (Merleau-Ponty, 2012, pp. 72-73).

For example, the convenience store across the street pre-exists my looking out at it from my window: I did not invent it out of whole-cloth in my mind. However, I can only *actually* perceive it by seeking to merge that pre-existing external reality with my own self: I see its red facade by erasing the distinction between my mental conception of the color and the physical reality of the facade's material (Merleau-Ponty, 2012, p. 69). Similarly, as I write this essay, I know that my keyboard is a differentiated, external object. But to use it, I seek to erase the boundary between the keys and my own fingertips: I stop having to consider the position of each key as a separate and external fact, and simply write.

In this way, Merleau-Ponty's phenomenology and Della Rocca's ascent deeply resemble each other: the boundary-erasing of Merleau-Ponty's perception is akin to Della Rocca's assertion that "in making the Parmenidean Ascent – and in seeing in this way – we literally are or become or are no longer distinct from the world [...]" (2020, p. 225). We can therefore put Merleau-Ponty's account in terms of ascents: to perceive and interact with the world we must constantly undertake a Parmenidean Ascent towards it.

That requirement of *constantly undertaking* an ascent, however, is an important divergence from Della Rocca's own conception. Della Rocca asserts that becoming indistinct from the world we see is a stable endpoint: once we have undertaken an ascent, we need not undertake it again. This goes against Merleau-Ponty's initial requirement, that to perceive an object it must pre-exist and be differentiated from our seeing of it.

However, I believe that Della Rocca's ascent is not free of this requirement. If Della Rocca's arguments are self-undermining, if they are "a relational metaphysical challenge to relational metaphysics itself" (Della Rocca, 2020, p. 220), they *must* presuppose differentiated being and the relational metaphysics inherent to such a conception. The only way in which *The Parmenidean Ascent* can make its claims is by acknowledging that they start from a *different place* than where they end. I here echo Michael A. Rosenthal's assertion that Della Rocca's monist account must lead us to "descend back into the midst of endless relations precisely to find a reason to accept the One" (2025, pp. 361–362).⁸

Such a move might be countered by asserting that though we may return to a *semblance* of differentiation, we do so while knowing that this appearance of differentiation holds no real metaphysical significance (Della Rocca, 2025, pp. 374–75). To do so, however, still renders Della Rocca's arguments inert: if the Principle of Sufficient Reason holds no significance, his call for us to follow it is also insignificant.

Though Della Rocca may accept this inert end, I would like to propose that there is value in the exact opposite: in embracing the continuous potential of his arguments and of the broader

⁸ It is important to note that Rosenthal's account and mine *do not* arrive at this conclusion through the same means, his being significantly more focused on the ethical possibilities of differentiation versus undifferentiation.

concept of making an ascent *instead* of the inert endpoint they attempt to bring about.

Living with Perpetual Motion

To exemplify what I mean by potential, take this very paper. If it is true that we must continuously ascend, then this essay is doomed to be incomplete: I must continuously seek to erase the boundary between this essay and the context it is present in, bringing in more and more readings, letting it continuously expand in scope. On the other hand, if I accepted that I could complete this ascent in writing, by erasing the boundary between this essay and everything else, this essay would be worthless: it would be indistinguishable from the world, the difference between its existence and non-existence invalidated. It is a state Hegel describes as “equality with itself, complete emptiness, absence of all determination and content” (2014, chap. 1, sec. B). To truly achieve completeness would be to destroy it and me both, which I would (perhaps arbitrarily) prefer not to happen.

But this incompleteness, this state of restlessness, of perpetual striving towards a non-reachable completion, need not be a curse. Indeed, the persistent feeling that there is *more* to be found should not scare us, but be a call. We may never be able to appreciate an eventual conclusion wherein the entire world is made clear and known to us, but we *can* always make more things clear.

Therefore, to value the constant ascent is to say that though our final goal is unachievable, there *is* value to be found in seeking it. I find this more compelling than the conception of monism as an end-state, of which as previously discussed even Della Rocca seems to have little to say.

This appreciation of restless striving is also not exclusive to the practice of philosophy itself (nor could it be, as it demands that we move to erase such boundaries between practices). One such resonance is in discussions of dependence: as described by disability studies scholar Shane Clifton (2018, p. 125), there is no clear boundary between independence and dependence; rather, they are in constant motion, as we constantly vary how much we rely on others. Rather, Clifton proposes that we think of the continuous process of interdependence, as we constantly both give to and receive from others.

Another case where we find appreciation rather than fear of motion is in Julianne N. Chung's (2023, pp. 134, 137, 139) writing on grief, which espouses the value of maintaining the potential found in our relationships with persons who are now deceased, instead of accepting a clear end-state. In fact, Chung (2023, p. 134) calls us to appreciate "the wandering that is constitutive of human life" as we spend continuous effort in not letting our emotions coalesce (be it in absolute sadness or absolute joy), instead embracing continuous creativity in striving between one and the other.

I highlight these examples to show that the inconclusiveness of *The Parmenidean Ascent* as viewed through a

dialectic lens is not a bug, but a feature, and one that we might quite happily live with.

(non-)conclusion

My aim with this essay is not to do away with Della Rocca's monism, nor with relational metaphysics. It is to do neither, and keep both in constant tension. There is insight in erasing distinctions, and so there must be distinctions to erase, and so we must erase them. In so doing, we can embrace, rather than extinguish, the potential of *The Parmenidean Ascent*. We must keep on striving.

And take your foot off of the brake
For Christ's sake!

- Dilaudid (2005), *The Mountain Goats*

References

- Casati, Filippo. 2025. "Della Rocca's Darkest Hour." *European Journal of Philosophy* 33 (1): 325–38.
<https://doi.org/10.1111/ejop.13051>.
- Chung, Julianne Nicole. 2023. "'See You in Your Next Life': Creativity, the Zhuangzi, and Grief." *Res Philosophica* 100 (1): 121–49. <https://doi.org/10.11612/resphil.2265>.
- Clifton, Shane. 2018. *Crippled Grace: Disability, Virtue Ethics, and the Good Life*. Baylor University Press.
<https://muse.jhu.edu/pub/170/monograph/book/57948>.
- Conant, James. 1992. "The Search for Logically Alien Thought: Descartes, Kant, Frege, and the Tractatus." *Philosophical Topics* 20 (1): 115–80.
- Diamond, Cora. 1991. *The Realistic Spirit: Wittgenstein, Philosophy, and the Mind*. The MIT Press.
<https://doi.org/10.7551/mitpress/5797.001.0001>.
- Della Rocca, Michael. 2020. *The Parmenidean Ascent*. Oxford University Press.
- Della Rocca, Michael. 2025. "Replies to Critics of The Parmenidean Ascent." *European Journal of Philosophy* 33 (1): 365–76.
<https://doi.org/10.1111/ejop.13054>.
- Hegel, G W F. 2014. *Science of Logic*. 1st ed. Routledge.
<https://doi.org/10.4324/9781315823546>.
- Merleau-Ponty, Maurice. 2012. *Phenomenology of Perception*. Translated by Donald A. Landes. Routledge.
- Rosenthal, Michael A. 2025. "Parmenides and Dr. Strangelove, Or How I Learned to Stop Worrying about Monism and Accept the World of Relations, at Least for the Sake of the Good." *European Journal of Philosophy* 33 (1): 354–64.
<https://doi.org/10.1111/ejop.13057>.
- Ware, Benjamin James. 2015. *Dialectic of the Ladder: Wittgenstein, the "Tractatus" and Modernism*. Bloomsbury Academic.

Dignity to Forests: Introducing the Notion of Honor towards Old-Growth Forests in the “War in the Woods”

The “war in the woods” was the first mainstream environmental movement in British Columbia, Canada, advocating the protection of old-growth forests from unsustainable logging practices prevalent in the province. The movement focused on protecting old-growth forests in Clayoquot Sound, BC, in response to the provincial government’s decision to allow the forestry company Macmillan Bloedel to clear-cut the area. Environmental groups such as the Western Canada Wilderness Committee (WCWC) and the Peace Camp shared visual and educational messages to change moral attitudes towards old-growth forests. The movement culminated in a blockade that garnered national and international media attention, leading to the suspension of logging in the area. Applying Kwame Anthony Appiah’s argument for the crucial role honor can play in moral revolutions to the “war in the woods” highlights how environmental groups took steps towards establishing the preservation of old-growth forests as a source of national honor and introduced to the mainstream the notion that these forests had dignity in attempts to transform moral attitudes in BC. Hence, the “war in the woods” protest provides a framework for appealing to the dignity of non-human beings.

The “War in the Woods” was the first mainstream environmental movement in British Columbia, Canada, advocating for the protection of old-growth forests from unsustainable logging practices that were prevalent in BC.¹ A significant portion of the movement focused on protecting old-growth forests in Clayoquot Sound, BC, in response to the provincial government’s decision to allow the forestry company Macmillan Bloedel to clear-cut the area.² Protestors formed a blockade that garnered national and international media attention, resulting in the suspension of logging in the area.³ Applying Kwame Anthony Appiah’s argument for the crucial role honor can play in moral revolutions to the “War in the Woods” reveals how environmental groups took steps towards establishing defending the preservation of old-growth forests as a source of national honor and introduced to the mainstream the notion that these forests had dignity in attempts to transform

¹ Nina Beveridge, producer, *The Nature of Things*, season 62, episode 10, “War for the Woods,” directed by Geoff Morrison and Sean Stiller, featuring David Suzuki, and Stephanie Kwetásel’wet Wood, aired March 17, 2023. CBC, 2023, Curio.

² Alia Dharssi, “Communication, social change, and the ‘war in the woods’: a study of how the Western Canada Wilderness Committee shaped the public’s understanding of the wilderness” (Graduating Essay (Hons. B.A.), University of British Columbia, 2009), 71.

³ Bruce Braun, *The Intemperate Rainforest: Nature, Culture, and Power on Canada’s West Coast* (University of Minnesota Press, 2002), 1, <https://muse.jhu.edu/book/31947>.

moral attitudes in BC.⁴ Hence, the “War in the Woods” protest provides a framework for appealing to the dignity of non-human beings.

This paper will first recount the “War in the Woods,” highlighting the Clayoquot Sound blockade. Then, it will discuss Appiah’s account of honor and examine how environmental groups that participated in the “War in the Woods” invoked *national honor*, a form of pride grounded in one’s affiliation to a region,⁵ in an attempt to instigate moral change in BC. Further, I will relay Appiah’s account of the function served by appeals to the dignity of working-class men during England’s abolition of slavery, and claim that protestors similarly invoked the dignity of old-growth forests to halt logging during the “War in the Woods.”⁶ Finally, I’ll consider some challenges Kimberlee Brownlee poses toward the success of a moral revolution regarding the environment,⁷ and highlight responses found in Bruce Braun’s study of Canadian West Coast

⁴ Kwame Anthony Appiah, *The Honor Code: How Moral Revolutions Happen* (W. W. Norton, 2011), 12, Google Play Books.

⁵ Appiah, *Honor Code*, 134–35.

⁶ Appiah, *Honor Code*, 154.

⁷ Kimberlee Brownlee, “On the Urgency of Kickstarting a Moral Revolution to Save Ourselves,” in *Conversations in Philosophy, Law, and Politics*, ed. Ruth Chang and Amia Srinivasan (Oxford, 2024), <https://doi.org/10.1093/oso/9780198864523.003.0001>.

rainforests,⁸ and Christopher D. Stone's paper "Should trees have standing? Toward legal rights for natural objects."⁹

The "War in the Woods": The Movement to Protect British Columbia's Old-Growth Forests

The "War in the Woods" in Clayoquot Sound, BC, began in 1993 with the provincial government's Clayoquot Decision,¹⁰ which permitted the logging of two-thirds of the area's old-growth forest by Macmillan Bloedel. BC defines old-growth forests as forests containing trees at least 140–250 years old, which the forestry sector values for their high wood yields. Indigenous peoples and environmentalists in BC immediately opposed the decision, since the method most forestry companies employ, of clear-cutting large areas of old-growth forest, negatively impacts the environment. Many protestors instead advocated for

⁸ Braun, *Intemperate Rainforest*.

⁹ Christopher D. Stone, "Should trees have standing? Toward legal rights for natural objects," *Journal of Human Rights and the Environment* 3 (2012): 4–55, <https://doi.org/10.4337/jhre.2012.02.02>.

¹⁰ Alia Dharssi, "Communication, social change, and the 'war in the woods': a study of how the Western Canada Wilderness Committee shaped the public's understanding of the wilderness" (Graduating Essay (Hons. B.A.), University of British Columbia, 2009), 70–71

selective logging, in which loggers cut specific trees rather than large areas of forest.¹¹

The provincial government was uninterested in compromising with the protestors, so they organized a blockade to contest the Clayoquot Decision.¹² Protestors gathered daily for three months on a secluded logging road, intervening vehicles delivering loggers to their worksites.¹³ The protestors successfully obstructed the logging trucks on August 9th, leading to the arrest of 300 participants and garnering international attention for the cause. The Clayoquot Sound blockade brought the issue of extensive logging of old-growth forests into the mainstream,¹⁴ producing international media recognition and generating compelling images for the environmental cause.¹⁵ As the summer progressed, protestors arrived internationally, including popular figures like Robert Kennedy Jr., Australian rock band Midnight Oil, and members of the Green Party from the European Parliament.¹⁶ A little under 10,000 people of all ages joined the blockades by the end of the

¹¹ Beveridge, *The Nature of Things*.

¹² *Ibid.*

¹³ Braun, *Intemperate Rainforest*, 1.

¹⁴ Beveridge, *The Nature of Things*.

¹⁵ Dharssi, "Communication, social change, 'war in the woods,'" 25.

¹⁶ Braun, *Intemperate Rainforest*, 1.

summer of 1993¹⁷ to oppose a deeply ingrained industry in BC,¹⁸ creating one of the largest acts of civil disobedience in Canadian history.¹⁹ Macmillan Bloedel postponed logging in the Clayoquot Sound area as national and international media attention and threats of consumer boycotts increased.²⁰ In 1998, the company agreed to transfer 51% of its rights to a First Nations logging company that committed to preserving certain areas.²¹ The Clayoquot Sound blockade was a precursor to an international boycott of Macmillan Bloedel paper products and several other anti-logging demonstrations.²² The protest's message questioned the method of forest management in BC and the body that administered it. It also challenged the economic structure instituted by the provincial government and industry following World War II by questioning the foundation of ideas of progress and resource liquidation underlying the province and its lumber industry.²³ BC's environmental movement

¹⁷ Pierre Walter, "Adult Learning in New Social Movements: Environmental Protest and the Struggle for the Clayoquot Sound Rainforest," *Adult Education Quarterly* 57, no.3 (2007): 248- 249, <https://doi.org/10.1177/0741713606297444>.

¹⁸ Beveridge, *The Nature of Things*.

¹⁹ Braun, *Intemperate Rainforest*, 2.

²⁰ *Ibid.* 2.

²¹ Dharssi, "Communication, social change, 'war in the woods,'" 71.

²² Walter, "Adult Learning," 254, 249.

²³ Dharssi, "Communication, social change, 'war in the woods,'" 1.

challenged the prevailing discourse of progress by questioning what delineates progress and whether it outweighs the cost.²⁴

The Role of Honor in the “War in the Woods”

Kwame Anthony Appiah’s *The Honor Code: How Moral Revolutions Happen* offers a lens for understanding the shift in moral attitudes during the “War in the Woods” by examining the role of honor in the protests. In *The Honor Code*, Appiah questions what one can learn about morality by exploring moral revolutions, which are radical changes in a society’s general moral attitude and coinciding practices.²⁴ He discusses three past and one contemporary moral revolution and explores the role of honor in each. Appiah employs Aristotle’s definition of ethics as the study of the best life to characterize the dimension encompassing morality, the guiding value that determines what one owes to others.²⁵ Appiah asserts that honor is particularly influential in changing deeply ingrained beliefs because it appeals to inescapable features of human social psychology.²⁶ Appiah’s key argument is that honor is especially effective in transforming private moral attitudes into public norms, which

²⁴ *Ibid.* 19.

²⁵ Appiah, *Honor Code*, 11.

²⁶ *Ibid.* 13–14.

he claims historians and sociologists rightly assert is crucial to the prosperity of such political movements.²⁷

Appiah defines an honorable person as one who aims not merely to be respected but to be worthy of respect.²⁸ Honor codes are standards of behavior and action that confer recognition respect and esteem, often while neglecting morality. One who complies with honor codes will respect those who adhere to them and feel contempt for those who do not. Therefore, people experience self-respect if they satisfy these standards, and self-contempt or shame if they do not.²⁹ An honor world is a collective that recognizes the same honor codes; it comprises all who apprehend and accept an honor code and hence transcends one's own society.³⁰ One's honor peers are those with the same identity and honor world.³¹

Appiah employs Stephen Darwall's two forms of respect in his account of honor: appraisal respect and recognition respect, each coinciding with a type of honor. Appraisal respect, also called esteem, involves positively evaluating someone according to a standard, where performing well on that standard requires doing better than most others. One gains esteem

²⁷ *Ibid.* 215.

²⁸ *Ibid.* 198.

²⁹ *Ibid.* 36.

³⁰ *Ibid.* 39, 107.

³¹ *Ibid.* 196.

through competitive honor, measured in degrees. One bestows recognition respect by regarding others in ways that sufficiently acknowledge some fact about them, rather than by achieving a standard.³² Recognition respect can coincide with a positive or negative perspective, but the type of respect required for honor always aligns with a positive one. One derives recognition respect from peer honor, which regulates relationships among equals. One does not receive peer honor in degrees; one either has it or does not. Appiah emphasizes that the differentiation between forms of honor is merely theoretical, as these variations are not invariably distinct in practice.³³

Applying Appiah's account of honor to the "War in the Woods" reveals how honor influenced BC citizens' moral attitudes towards old-growth forests. Environmentalist groups like the Western Canada Wilderness Committee (WCWC) and the Peace Camp³⁴ achieved this shift by challenging the predominant discourse of progress that prioritizes economic advancement over the preservation of old-growth forests. They achieved this by disseminating visual and educational messages in Clayoquot Sound and through the media throughout the

³² *Ibid.* 33, 195.

³³ *Ibid.* 33.

³⁴ Walter, "Adult Learning," 260.

“War in the Woods,” which framed protecting old-growth forests as part of the honor code of a BC citizen.³⁵

Paul George and Richard Krieger established the WCWC in 1980, at the beginning of the “War in the Woods.” By 1991, the WCWC had 15,000 members and 20,000 other unofficial supporters.³⁶ The committee regarded educating the public and changing its perspective towards wilderness as key to accomplishing their aims.³⁷ Thus, they disseminated their message of wilderness conservation to the public through images of BC’s natural landscape and educational approaches, aiming to mold and animate public sentiment to demand politicians’ attention. The WCWC led citizen mobilization and indirect lobbying in the BC wilderness conservation campaign.³⁸ The committee contested logging old-growth forests like Clayoquot Sound, but they refused to partake in the blockade, upholding a strong commitment against engaging in civil disobedience.³⁹

To spread their message of conserving old-growth forests during the “War in the Woods,” the WCWC produced several publications and campaigns intended to shape how people

³⁵ Dharssi, “Communication, social change, ‘war in the woods,’” 7, 17, 46.

³⁶ *Ibid.* 2.

³⁷ *Ibid.* 43.

³⁸ *Ibid.* 2–3.

³⁹ *Ibid.* 20, 25.

construct their concept of wilderness within the BC economy and society. One of these publications was a coffee table book titled *Clayoquot: On the Wild Side*, published in 1990.⁴⁰ The book includes photographs of Clayoquot Sound's old-growth forests and text advocating for their preservation. In the art and text, the WCWC frames the wilderness as a revered escape from civilization that must remain in its initial, pristine form. The book describes Clayoquot Sound as an "evolutionary masterpiece" as "old as the classical ruins of Greece and Rome []" thereby bestowing substantial value on the site within Western culture by likening it to revered Western monuments. The publication also evoked a sense of imminent loss, compelling Canadian society to protect the remaining significant wilderness areas as representations of Western heritage. By framing Clayoquot Sound's old-growth forests as meaningful to Western heritage, the WCWC's publication characterized these sites as sources of national honor; therefore, to protect them is to protect the honor of one's nation. Hence, such publications helped motivate citizens to protect old-growth forests by framing such behavior as defending the honor conferred upon them as members of their nation.⁴¹

⁴⁰ *Ibid.* 8.

⁴¹ *Ibid.* 6.

The committee also harnessed the mass media to increase public recognition of the environment.⁴² The positive response among BC's middle and upper classes to images disseminated by the WCWC in their publications and the media was critical to amassing public endorsement and prompting government response.⁴³ The WCWC often employed the press to publicize the trails they built, using them to garner support for preservation movements by portraying these trails in ways that supported wilderness conservation.⁴⁴ While trail construction did not directly impact the "War in the Woods" conflict, it inspired the public and media to adopt a preservation-based perspective of the environment.⁴⁵ Former Canadian Liberal environmental minister Charles Caccia expressed this conservation-oriented attitude during an interview with the Vancouver Sun while he was visiting one of the WCWC's trails, claiming BC's forestry operations were "barbaric" and "not worthy of a civilized society such as ours."⁴⁶ Caccia appealed to national honor, asserting that disregard for the preservation of old-growth forests is not respectable in Canadian society and thus dishonorable.

⁴² *Ibid.* 4.

⁴³ *Ibid.* 18.

⁴⁴ *Ibid.* 4, 25, 29, 33.

⁴⁵ *Ibid.* 37-38.

⁴⁶ *Ibid.* 28.

Regarding which of Appiah's types of honor environmental movements invoked during the "War in the Woods," people may have appealed to competitive honor because those who advocated for old-growth forest preservation more publicly may have received more respect. However, it appears that most organizations invoked peer honor, as best exemplified by the Peace Camp. The Peace Camp was a collective that participated in the Clayoquot Sound blockade and was key to facilitating individual and group learning and change among protestors and the broader public.⁴⁷ The Peace Camp drew on the history of civil disobedience in past feminist movements, sharing a dedicated commitment to feminist principles of non-violence and consensus, which earned the group high public regard and influenced informal supporters to become protestors. The Peace Camp maintained three central rules: "equality, non-violence, and decisions by consensus." Protest leaders espoused these rules daily, and they were embedded in the method and construction of the Peace Camp and the blockade, and reiterated when new protestors joined. The strength of the Peace Camp's commitment to these rules indicates that the extension of honor for acting in accordance with the code of supporting old-growth forest preservation would be equal rather than contingent on

⁴⁷ Walter, "Adult Learning," 260.

performance.⁴⁸ Protestor Laurie Brant, an interviewee in the film *Fury for the Sound: The Women at Clayoquot*, exemplified the appeal to peer honor, saying, “If I can’t [defend the preservation of BC’s old growth forests], what kind of a parent am I? What kind of a citizen am I? What kind of a mother am I?”⁴⁹ Brant’s statement exemplifies the view that defending old-growth forests confers honor or shame on a BC citizen.

Conferring Dignity upon Trees

Another facet of Appiah’s theory of honor that is applicable to the “War in the Woods” is dignity, as protestors appealed to the notion of conferring dignity on trees. Appiah discusses the function of dignity when analyzing the moral revolution of the British abolition of slavery in 1833. Appiah describes slavery as the subjugation of one race to another, which necessitated the structural subjection of Black people to dishonor. The abolition of slavery conflicted with British economic interests, as did halting logging in BC’s Clayoquot Sound.⁵⁰ Nonetheless, English citizens eventually collectively condemned slavery, thereby using their denunciation as a source of national honor.⁵¹

⁴⁸ *Ibid.* 254–55.

⁴⁹ *Ibid.* 257.

⁵⁰ Appiah, *Honor Code*, 124–26.

⁵¹ *Ibid.* 153.

However, this was not the only British appeal to honor that led to the abolition of slavery. As the British working class grew a collective consciousness, they formulated a stance against slavery, because it signified that the labor which defined them was dishonorable.⁵² Therefore, the unified British working class's stance against slavery linked them to a new emblematic investment in their dignity.⁵³

As the English working class unified during the democratic age, the concept of dignity emerged as people began to inquire how to apply the notion of honor, considering the modern revelation that all humans are essentially equal under moral law. The contemporary standard of equality began with the idea that specific grounds for treating people unequally are not legitimate before determining what is. Thus, discrimination, in the sense of discrepancies in the treatment of others grounded in their social identities, began to require explanation. Following the French and American revolutions, people began to dispute the view that one's social class was a legitimate basis for discrimination. Over the last century, society has directed the same realization toward race and gender. The Universal Declaration of Human Rights, adopted in 1948, states that dignity positively demands equality. Appiah defines dignity as the "right to respect people have simply in virtue of their

⁵² *Ibid.* 144.

⁵³ *Ibid.* 154.

humanity.” Facts considered in acknowledging human dignity, or the “grounds of dignity,” justify regarding people in ways that acknowledge such essential obligations and abilities. These grounds are that one has the potential to create a meaningful life, the capacity to suffer, love, and create, and they require sustenance, shelter, and recognition from others. Hence, dignity entails the right to recognition respect, since one obtains respect in virtue of specific facts about them. Although dignity is unearned, one loses it by refusing to behave in a way that is consistent with one’s humanity. Ultimately, while appeals to British national honor contributed to the English abolition of slavery, the invocation of working-class dignity profoundly influenced this moral shift.⁵⁴

Appiah argues that in a society like that of England before the abolition of slavery, which withheld respect from the working class, demanding their dignity was a radical proposal.⁵⁵ As conferring dignity upon working people appeared radical then, conferring dignity upon old-growth forests appears radical now. Nonetheless, protestors and environmental groups participating in the “War in the Woods” seemed to invoke the idea that old-growth forests have dignity. A potential challenge to the proposal of an appeal to dignity for old-growth forests is that, despite their adherence to Appiah’s other conditions, he

⁵⁴ *Ibid.* 147–51.

⁵⁵ *Ibid.* 151.

defines dignity as necessarily associated with one's humanity.⁵⁶ However, in Bruce Braun's analysis of the "War in the Woods," he asserts that nature is socially constructed, challenging the strict regulation of boundaries between humans and nature. Braun claims society generates concepts of nature because what people regard as "natural" establishes ecological and social relationships.⁵⁷ Thus, one's perspective on nature does not merely mirror reality but constructs it, indicating that the notion of nature as socially established carries epistemological and ontological implications. The many historical and discursive processes that regulate how people experience and apprehend nature shapes their perception of it, and the naturalization of a "true construction" of nature at any given time is an effect of power that serves particular interests, like those of capital.⁵⁸ The rationale of global capitalism considers certain environments and communities disposable to the nation to obtain profit and progress. Therefore, nature in capitalist societies is a store of resources exploitable for wealth and nation-building.⁵⁹ The industrial forestry industry gains authority through discursive displacements and formulations based in modern capitalism that infuse entities with significance, creating

⁵⁶ *Ibid.* 149.

⁵⁷ Braun, *Intemperate Rainforest*, 11.

⁵⁸ *Ibid.* 262.

⁵⁹ *Ibid.* 2.

a hegemonic standard construction of the forest based on value and production, which many BC citizens perceive as essential to the province's welfare.⁶⁰ This hegemonic normalization of the forest as a profit source objectifies it, reinforcing a distinction between humans and nature that leads humans to refuse to confer traits perceived as essentially human onto nature. The idea that old-growth forests and other natural entities cannot attain dignity stems from this distinction. Hence, Braun advocates for an understanding of the forest as a culturally and politically embedded web of relationships, arguing that referring to forests in this way highlights the erasure of other perspectives of nature that defy defining it in terms of resource extraction.⁶¹ Additionally, he asserts that because human and nonhuman activity equally construct nature, it is hybrid, and thus one should not strictly enforce distinctions between humans and nature. The concept of nature as a connected web that simultaneously incorporates humans, animals, and machines prioritizes one's responsibility for these connections rather than imposing regulatory boundaries. By perceiving nature as a web of relations encompassing humans, the boundaries between humans and nature fade,⁶² eroding the obstacles to conferring dignity upon trees, as it ceases to be a

⁶⁰ *Ibid.* 249.

⁶¹ *Ibid.* 260.

⁶² *Ibid.* 13.

distinctly human trait. The WCWC and Peace Camp treated old-growth forests with dignity by advocating for their protection as beings with lives as significant as those of humans throughout their publications and educational campaigns. Each of the WCWC's visual publications defies the logic of nature as a meaningless source of capital by infusing it with importance within the Western cultural context. The WCWC's publications situated old-growth forests as sites of cultural significance in the Western context by accompanying photographs of these landscapes with textual descriptions that likened the forests to established monuments in Western culture. One such publication about clear-cutting in BC's Carmanah Valley records the testimony of participant Cameron Young, who claimed, "[to] clear-cut the Carmanah rainforest would be to desecrate a global heritage—the equivalent of bulldozing the Parthenon."⁶³ Young's statement exemplifies how comparing old-growth forests to Western cultural monuments justifies their protection as important to Western heritage.

The Peace Camp transmitted their view of equality, which was key to their educational and demonstrational systems, to old-growth forests, thereby imbuing them with dignity.⁶⁴ The Peace Camp achieved this by espousing narratives that highlighted the inherent value of old-growth forests through

⁶³ Dharssi, "Communication, social change, 'War in the Woods,'" 55.

⁶⁴ Walter, "Adult Learning," 255.

various practices and spectacles.⁶⁵ One such event was a service held for fallen trees on Remembrance Day.⁶⁶ By mourning trees in the same manner as soldiers, the Peace Camp glorified these trees as having experienced lives of equal significance to soldiers, thereby conferring the same amount of honor onto both parties. The service also reinforced the perspective that trees deserve recognition through ceremony. Therefore, the practices of the Peace Camp demonstrated that trees possess dignity and hence deserve respectful treatment. Similar to how English society abolished slavery after recognizing that the working class and Black enslaved people deserved dignity,⁶⁷ BC citizens decided to protect the old-growth forests of Clayoquot Sound because environmental groups appealed to the idea of trees as subjects with dignity.

Although environmental groups successfully appealed to honor to protect old-growth forests during the “War in the Woods,” philosopher Kimberlee Brownlee questions whether honor can initiate a moral revolution regarding the non-human world. While Brownlee applies concepts of honor to an environmental revolution aimed at addressing the climate crisis, two of her objections apply to attitudes towards old-growth forests. First is the humanistic assumption, which questions

⁶⁵ *Ibid.* 259.

⁶⁶ *Ibid.* 256.

⁶⁷ Appiah, *Honor Code*, 154.

whether humans have innate value and complete moral status. Brownlee argues that, because of the prominence of speciesism and personism, which frame humans as superior to other organisms, humans should question whether their attributes are sufficient to grant moral status, especially if they maintain that posing an inescapable threat reduces someone's moral standing.⁶⁸ However, adopting Braun's view of forests as webs of social connection, wherein humans and nature are not strictly distinct, responds to Brownlee's challenge.⁶⁹ If one does not perceive humans as distinct from nature, there is no difference in the threat they pose to each other, and hence one entity can dictate the moral status of another.

The second relevant challenge Brownlee poses is the possibility assumption, which questions whether humans can trigger moral revolutions in their treatment of non-human beings.⁷⁰ Brownlee asserts that human supremacy is an inevitable facet of human moral psychology, highlighting that each of Appiah's cases of moral revolution discussed in *The Honor Code* concerns the treatment of humans.⁷¹ Braun addresses

⁶⁸ Brownlee, "On the Urgency of Kickstarting a Moral Revolution to Save Ourselves," 6-8.

⁶⁹ Braun, *Intemperate Rainforest*, 13.

⁷⁰ Brownlee, "On the Urgency of Kickstarting a Moral Revolution to Save Ourselves," 4.

⁷¹ *Ibid.* 13-14.

this challenge by claiming that while it is nearly impossible to maintain a relationship that is not instrumental in some regard, this does not entail that humans cannot value nature beyond its utility.⁷² Legal professional Christopher Stone supported this perspective, arguing that because human prosperity depends on the environment's well-being, these aims are frequently mutually reinforcing, making it difficult to distinguish whether appeals to collective interests include those of the environment.⁷³ Therefore, despite human supremacy, humans can value nature beyond its profitability and pursue goals that enhance its well-being.

Conclusion

Ultimately, applying Appiah's account of the operation of honor in moral revolutions to the "War in the Woods" reveals how environmental groups appealed to honor in an attempt to transform BC citizens' attitudes towards the preservation of the province's old-growth forests. The WCWC and Peace Camp, environmental collectives that were influential in the conflict, espoused a new narrative that permitting the irresponsible destruction of old-growth forests dishonors one as a BC citizen, and produced visuals and symbols that appealed to the notion

⁷² Braun, *Intemperate Rainforest*, 257.

⁷³ Stone, "Should trees have standing?" 43.

that these forests have dignity, justifying their protection. Bruce Braun's view of nature as a social web of relations defends the idea of trees having dignity, which is predominantly framed as a human trait, by diminishing the importance of barriers between humans and nature.⁷⁴ The moral revolution that occurred during the "War in the Woods" demonstrates that honor need not be restricted to humans, or even to old-growth forests. The environmental group's appeals to the honor of non-human beings offer a path to conferring honor on all non-human living creatures.

⁷⁴ Braun, *Intemperate Rainforest*, 13.

References

- Appiah, Anthony Kwame. *The Honor Code: How Moral Revolutions Happen*. W. W. Norton, 2011. Google Play Books.
- Beveridge, Nina, producer. *The Nature of Things*. Season 62, episode 10, "War for the Woods." Directed by Geoff Morrison and Sean Stiller, featuring David Suzuki, and Stephanie Kwetásel'wet Wood. Aired March 17, 2023. CBC, 2023, Curio.
- Braun, Bruce. *The Intemperate Rainforest: Nature, Culture, and Power on Canada's West Coast*. University of Minnesota Press, 2002. <https://muse.jhu.edu/book/31947>.
- Brownlee, Kimberley. "On the Urgency of Kickstarting a Moral Revolution to Save Ourselves." In *Conversations in Philosophy, Law, and Politics*, edited by Ruth Chang and Amia Srinivasan. Oxford, 2024. <https://doi.org/10.1093/oso/9780198864523.003.0001>.
- Dharssi, Alia. "Communication, social change, and the 'War in the Woods': a study of how the Western Canada Wilderness Committee shaped the public's understanding of the wilderness." Graduating Essay (Hons. B.A.), University of British Columbia, 2009.
- Nash, Roderick Frazier. "A Wilderness Condition." In *Wilderness and the American Mind*. Yale University Press, 2014. Google Play Books.
- Stone, Christopher D. "Should trees have standing? Toward legal rights for natural objects." *Journal of Human Rights and the Environment* 3 (2012): 4-55. <https://doi.org/10.4337/jhre.2012.02.02>.
- Walter, Pierre. "Adult Learning in New Social Movements: Environmental Protest and the Struggle for the Clayoquot Sound Rainforest." *Adult Education Quarterly* 57, no. 3

(2007): 248-263.

<https://doi.org/10.1177/0741713606297444>.

VTubers: Para-Social Relationships with Simulacra on Virtual Hyperreality

VTuber is a phenomenon that has gained popularity in the last ten years over the internet. They are fictional personas embodied by digital graphics played by an actor, hosting interactive live streams involving bi-directional interaction with thousands of viewers. Many of the viewers developed para-social relationships with their favourite VTubers through watching and interacting with them through the digital media. This paper examines the VTuber phenomenon through a reading of Baudrillard's Simulacra and Simulation, which introduced the concept of hyperreality as a generation by models of a real without being rooted on profound reality. Two main arguments discussed are: that the viewer-VTuber relationship is between the viewer and The Person Behind as individuated from both the fictional character and the actor, and that The Person Behind is hyperreal. A deceptive effect brought up by the VTubers as simulacra is examined through Baudrillard's example of a hyperreality. The hyperreality of VTubers opens up the possibility for the simulacra to precede the real.

Kizuna AI, the first VTuber, debuted in 2016 as a virtual YouTuber, hence the term VTuber (Turner, 2022). Kizuna AI is a digital persona represented by a digitally modeled character with an actor behind it, controlling her movements, facial

expressions, and voice. The term VTuber has broadened to refer to any virtual entity that posts digital content and hosts live streams on the internet regardless of the platform, similar to what traditional live-streamers do.

A VTuber embodies a fictional digital persona in the form of digital graphics. It is a character that usually comes with a background story, or at least, a concept of what it should be perceived as. The background story of a VTuber is not aimed to make sense of the character in their fictional world, but merely to form an identity so they can stand out. It could be as bizarre as “a shark goddess from Atlantis that takes the form of a human with sharp teeth” or as simple as “a detective girl”. A VTuber involves a concept of identity, digital representations, and an actor behind it. However, there are important distinctions in how the actor influences the character of the VTuber compared to other media such as film. In film, the actors are supposed to hide their real self as much as possible in order to play the character. Meanwhile, the actor behind a VTuber often shapes the character played with their own personalities and traits. This is apparent when VTubers talk about their favourite food, favourite games, their preference in music, their beliefs, etc. These dispositions might come from the actor and not part of the designed character. The viewers are fully aware of this and do not see this as a problem, unlike in film. In fact, a lot of the viewers are interested in connecting with the ‘actor’ via the VTuber character. The dynamic of the VTuber persona and the actor behind it will be further examined in section 2.

One characteristic for VTubers is that the viewing experience is interactive, their live streams through the internet allow viewers to interact with the VTuber via text in a chat pool. A number of viewers, sometimes tens of thousands, bombard the chat with their text submission hoping it would catch the eye of the VTuber. Viewers can get a better chance to be noticed by sending donations along with their messages. VTubers participate in a variety of different activities in their streams. They could be gaming, drawing, or simply chatting, the only limitation is that they must still be able to interact with the viewers.

Viewers watch and interact with their favourite VTuber as a stress relief on a regular basis (Tan, 2023). Although engagement with viewers is important to maintain the community and interactive facets of VTubing, some viewers simply watch the streams without interacting. It is as if the other viewers participating in the chat are already filling the community role for them. Other than the interactivity, seeing and hearing the VTubers talk is an attractive enough form of entertainment exemplified by the many video clips of VTuber streams on the internet that attract viewers even without the bi-directional interactions.

The interactions between viewers and VTubers on live streams can be considered para-social, which is a term that has been used since the 1950s. One of the prominent examples of the term at the time was the interaction between radio broadcasters and their listeners (Horton & Wohl, 1956). Radio listeners could interact with broadcasters through letters or phone calls, which

is comparable to the interactivity of a VTuber stream's chat. Para-social interactions are only in the context of mass communication. It differs from one-to-one communication because although the viewer-to-broadcaster interactions are one-to-one, the broadcast is, by definition, one-to-many, and the interactions are many-to-one.

These para-social interactions with a VTuber can result in what is called para-social relationships. It is, although tempting to say otherwise, a relationship that is neither social nor imaginary (Giles, 2010). It is not a social relationship because it does not necessarily involve direct communication between two parties with full awareness of each other as individuals. The VTuber does not intend to distinguish a viewer from another and build relationships with each of them individually. Moreover, viewers are usually only distinguishable from each other by their username. However, from the point of view of the viewers, the relationship is meaningful. However, the reality of the relationship and who is involved is a subject to philosophical inquiry. The following sections will examine the VTuber phenomenon and the ontology of it, and reveal these two points: that a viewer-VTuber relationship is between the viewer and *The Person Behind*, and that *The Person Behind* is *hyperreal*.

Ontological Examination of the VTuber

Giles (2010) acknowledged Busselle and Bilandzic's (2008) two types of realism, external and narrative realism, in the context of fictionality. External realism looks at how fiction makes sense in the real world, while narrative realism looks at the coherence and plausibility of the fiction in itself. This paper, however, will

refer to the *real* (Italicized) in the context of Baudrillard's simulacra – the *real* as distinct to the *hyperreal*. The *real* as in: how something is of profound reality. It is similar in spirit with Busselle and Bilandzic's external realism, but instead of dealing with the content of a fictional narration, it deals with the ontology of objects.

VTubers often discuss in-real-life experiences, and in doing so, they refer to themselves ambiguously. Consider this sentence uttered by a VTuber on a live-stream: "I went to the bookshop the other day, and got the book you guys recommended! I love the cover design!". Who is "I" in that sentence? We know that the fictional persona obviously did not go to a bookshop for it is a mere fictional character. We may intuitively think that it refers to the real actor behind the persona. It is plausible that the actual actor went to the bookshop but instead, I argue that the "I" here refers to another persona beside the fictional persona that the actor plays. In this paper, I will use the term *The Person Behind* to refer to this persona of the person behind the VTuber as experienced by the viewers and as either intentionally or unintentionally put forward by the actor. It does not refer to the actual actor behind. In this section, I will examine *The Person Behind* and support the argument that the counterpart in the para-social relationship of the viewer is *The Person Behind* as opposed to the *Vtuber persona* or the *Actor alone*.

How is *The Person Behind* individuated from the actual actor? The VTuber is a synthesis of the fictional persona and the actual actor. The amalgamation between the two generates *The Person Behind* as a consequence, which then becomes inseparable

from the VTuber. The VTuber is then built on the complex causal relation between the fictional persona, the actor, and *The Person Behind*. The viewers never interact with the actual actor outside the context of the VTuber. The real actor bears no necessity to the phenomenology of the VTuber as a whole, and therefore is replaceable. Suppose that at some point, the actor behind is replaced by another person who has learnt everything there is to know to play the character exactly as how the previous person did. The new actor is aware and able to replicate the preferences, psychological tendencies, way of speaking, voice characteristic via digital voice-changer, and so on. The viewers would be able to continue the para-relationship without any problem. Suppose that the viewers recommended the book before the change of actor, and the new actor was the one who went to the bookstore. The actor changed, yet *The Person Behind* persisted. *The Person Behind* was the one who received the recommendation of the book, and also the one who loves the cover design. The viewer interacted with *The Person Behind*, before and after the actor-change. The existence of *The Person Behind* is not between the VTuber persona and the actor, but the VTuber persona and the meta-awareness of the actor, which results in *The Person Behind*.

The actor-change is an exaggerated scenario to reveal the individuation of The Person Behind, but The Person Behind is individuated even without any change. When one watches a VTuber live-stream, they are well aware of the existence of The Person Behind, but might not realize the individuation of it from the actual actor. Section 3 will reveal the mechanism that plays

into the deceptive effect that prevents viewers to notice this individuation.

Another scenario that challenges the individuation of The Person Behind, and emphasizes the significance of the real actor, is the following. Suppose a VTuber is a member of a VTuber company, and at one point, the actor of that VTuber decides to leave the company for some reason. The actor then created her own fictional persona, and started her own VTuber channel. It is then made obvious, either by design or by chance, that the same actor that left the company is the one behind the new VTuber. The viewers acknowledge this, and some of the fans of the old VTuber transferred over to the new one, continuing their para-social relationship. It is thus very tempting to assume that the para-social relationship must be between the viewer and the actual actor. However, I argue that most people fail to see that The Person Behind is the one who persisted between both VTubers. The counterpart of the para-social relationship of the viewer is The Person Behind. It is true that The Person Behind has a tight causal relation with the actual actor, yet in both the old and new VTuber, the viewers do not engage with the actor directly, only via the VTuber persona, which is fictional and not grounded in profound reality. Moreover, the actor is still acting to put forward the persona of The Person Behind. This scenario, to the contrary of normative assumptions, actually plays a role of a simulacrum that precedes the real, hence creating a *hyperreality*.

Before the examination of *hyperreality* in the next section, it is helpful to understand that there is a social aspect

between different VTubers, where they interact with each other, and consequently, creating a complex web of causal relations enriching the fictional narrative. VTubers often hold collaborative live-streams where they hold conversations on many topics. In these interactions, the personality of The Person Behind is at play. VTubers put personal opinions and preferences into the table, which then influences the other VTubers. They bond and create meaningful friendships that viewers witness. These friendships are part of the fictional narrative, and so are all other relations. As another consequence, a collective memory of past collaborative events are formed. VTubers share memories of past events and also their perspective on them. They are part of this virtual social microcosm involving the Persons Behind. This is the basis of continuity of narrative between live-streams. It is how a friendship between two VTubers develops. It is worth noting that despite how the between-VTubers relationships are seen through the livestreams, they may not correspond to the relationships between the actors. This disparity further emphasizes the distinction between the social relations in profound reality and the VTubers' social microcosm – the former is of the *real*, while the latter is a simulation of the *hyperreal*.

The Hyperreality

Baudrillard (1994) picked Disneyland in America as a good example that involves all orders of simulacra. Simulacra are

objects of a simulation as images with varying degrees of relation with reality:

Such would be the successive phases of the image:
it is the reflection of a profound reality;
it masks and denatures a profound reality;
it masks the absence of a profound reality;
it has no relation to any reality whatsoever: it is its own
pure simulacrum.
(Baudrillard, 1994, p.6)

Disneyland is a potent example of a simulacrum as it provides characters played by actors that create a social microcosm full of emotion. Baudrillard went further and argued that Disneyland presents itself as imaginary in order to make people believe that the world outside its doors is a profound reality in contrast to inside, whereas profound reality is absent in the hyperreality outside its doors.

I argue that VTubers are pure simulacra even when they seem to allow a lot of bits and pieces of the reality of the actor to penetrate through the avatar. Nonetheless, they are simulacra that give rise to a social microcosm where virtual beings interact with each other while being watched by the viewers, much like a Disneyland park. Viewers are well aware that they are watching a simulated reality, a social dimension where VTubers have meaningful relationships with each other. Since the characters themselves are not real, the relationships between

them and the further story that develops through these interactions are just as unreal.

The awareness of *The Person Behind* their favourite VTuber gives a way for watchers to develop a parasocial relationship with the actor along with the avatar. The VTubers themselves often exploit this awareness to let loose and 'have a break' from all the acting. They deliberately discuss meeting each other in real life, talk about their impression on each other's actual appearance, and so on. It provides convenience for the VTubers because they can pick and choose when to act and alter how much effort they put into the acting. However, I argue that regardless of the degree of acting they put into it, they are nonetheless playing a persona as 'The Person Behind the VTuber' from the point of view of the viewers as subjects of hyperreality.

Instead of a back-and-forth between their real profound self and their avatar character, they are simply switching between two characters that are expressed only through simulacra. I argue that this back-and-forth between The Person Behind and the VTuber's persona plays the same effect as the gates in Disneyland. It is a distinction that tricks the viewers into believing the reality of one must be in contrast to the other.

Visitors enter Disneyland with the expectation of an experience of a magical social microcosm filled with joy, that emphasizes the contrast between it and the outside world. Being inside the apparent simulated environment triggers the belief that what is outside the simulation is 'real'. The same effect is played out by VTubers, by going back-and-forth between

playing the VTuber persona and the persona of The Person Behind, the contrast of reality between the two is emphasized. They make the viewers believe that the real person behind is let out through the avatar, while it is just a character behind another character. They make the viewers think 'If the VTuber persona is not the real persona of the subject, then The Person Behind must be the real one'. This effect of separation between realities is precisely what is found in Disneyland as Baudrillard (1994) wrote. The Person Behind, which is the actor's persona that is perceived as the 'real' person behind the avatar, is thus *hyperreal*. The viewers blindly accept this because the actual truth is irrelevant to the consumption of VTuber content as entertainment. They simply play along with the narrative of both simulacra of the avatar and The Person Behind.

As an example, consider a VTuber Abby, who here-and-there on her livestreams mentions 'in-real-life' stories that she is actually Canadian, unlike her VTuber persona that is an Atlantian goddess. Abby is close with her VTuber partner, Brenda, an Amazonian VTuber persona who is well aware that Abby is Canadian 'in-real-life', and reciprocally, Abby is also well aware that Brenda is 'in-real-life' Norwegian, not Amazonian. They often go on live streams together and talk about their 'in-real-life' identities. At this point, the viewers are brought into the hyperreality where *The Person Behind* Abby is Canadian, and *The Person Behind* Brenda is Norwegian. However, the unseen and unproven images of *The Person Behind* Abby and Brenda are pure simulacra. The viewers simply take them in as true. They naively play along with the statements of

where the persons behind are from. The Canadian person behind Abby and the Norwegian person behind Brenda are pure simulacra because they only exist in relation to their VTuber persona and all information about them is spoken through simulacra that are their avatars. We see the simulation in action when they interact with other VTubers, and the hyperreal information about each other's actor persona are reamplified throughout those interactions and therefore realized in the hyperreality.

Moreover, the viewers only engage with the persons behind through the simulacra. The truth behind their 'in-real-life' nationalities are irrelevant to the simulation perceived by the viewers. Whether or not the real persons behind are Canadian and Norwegian, does not change the experience of the simulated hyperreality. Also, even when they hold 'in-real-life' collaborations, they are in the context of 'the person behind the VTuber' and not somebody in profound reality. To further reveal the Baudrillardian simulation in this phenomenon, I can expose that there is no *real* person in the eyes of the viewer – only The Person Behind, which is hyperreal.

Baudrillard (1994) introduced the hyperreal as the generation by models of a real without origin or reality. It is a synthesis of simulacra that can precede the real. Pragmatically, it may seem that the causal relation between the actor and The Person Behind is linear, that is, the actor shapes The Person Behind. However, The Person Behind, as a simulacrum in a web of relation involving other VTubers as simulacra, generates the

hyperreal. This social microcosm of simulacra is precisely the *hyperreal*. It can precede the *real*.

The *hyperreal* precedes the *real* in ways that The Person Behind the VTubers influences the real actors to behave a certain way, and as a consequence, further develop the narrative. The *hyperreal* is no longer an image of any profound reality due to this cyclical causal relations. Moreover, the viewers are also engaged with the *hyperreal* through the interactions with the VTubers. They have a causal role in shaping the narrative, as instantiated with the book-recommendation example in S2.

Conclusion

We have established that the viewer-VTuber para-social relationship is between the viewer and The Person Behind, and that The Person Behind is *hyperreal*. The conclusion of the two premises is that the viewer-VTuber relationship is between the *real* and the *hyperreal*. This conclusion is under the claim that the viewers are *real* and not already *hyperreal* to begin with. However, there are strong pieces of evidence that ground the viewers in profound reality. First, the viewers are psychologically engaged despite engaging behind their anonymous username with the VTubers, as supported by the findings on stress-relief in Tan (2023). There are *real* effects that motivate the engagement and the development of the relationship, which is enough to show how the viewer is of profound reality. Second, many fanatic viewers gather in fan communities and meet each other in real life. The para-social

relationship with the *hyperreal* shapes social relationships of the *real*, that is, between the viewers. This conclusion leads to further philosophical inquiry on how a relationship of the *real* and the *hyperreal* may unfold, and what the consequences of these relationships may be. The awareness of *hyperreality* and its deceptive effect may potentially prevent the *real* from being eclipsed or eroded by the *hyperreal*.

References

- Baudrillard, J. (1994). *Simulacra and simulation* by Jean Baudrillard ; translated by Sheila Faria Glaser. The University of Michigan Press.
- Busselle, R., & Bilandzic, H. (2008). Fictionality and Perceived Realism in Experiencing Stories: A Model of Narrative Comprehension and Engagement. *Communication Theory*, 18(2), 255–280. <https://doi.org/10.1111/j.1468-2885.2008.00322.x>
- Giles, D. C. (2010). Parasocial Relationships. *Characters in Fictional Worlds*, 442–458. <https://doi.org/10.1515/9783110232424.4.442>
- Horton, D., & Wohl, R. (1956). Mass Communication and Para-Social Interaction: Observations on Intimacy at a Distance. *Psychiatry*, 19(3), 215–229. <https://doi.org/10.1080/00332747.1956.11023049>
- Tan, Y. (2023). More Attached, Less Stressed: Viewers' Parasocial Attachment to Virtual Youtubers and Its Influence on the Stress of Viewers During the COVID-19 Pandemic. *SHS Web of Conferences*, 155, 03012. <https://doi.org/10.1051/shsconf/202315503012>
- Turner, A. B. (2022). *Streaming as a Virtual Being : The Complex Relationship Between VTubers and Identity* (Dissertation). Retrieved from <https://urn.kb.se/resolve?urn=urn:nbn:se:mau:diva-53529>

From Reasons to Reason-Shaped Prose

Correctness depends on score.

– David Lewis, *Scorekeeping in a Language Game*

Large language models now routinely generate polished “explanations” of their outputs, and in practice we often treat these as if they were reasons. On Brandom’s picture, giving a reason is taking up a deontic status. You undertake commitments, you can lose entitlement when incompatibilities appear, and you are bound by how you repair or withdraw what you have said. I use that standard to ask a narrow question: when an LLM “explains” itself, is it actually playing that game, or only producing reason-shaped prose? I take one short answer from Claude Sonnet to the prompt “Should governments ban AI boyfriends and girlfriends?” and treat it as a compact policy text. I read its initial stance as a bundle of commitments, then apply two probes. First, I force a vocabulary shift toward morally loaded predicates. Second, I stipulate an empirical finding that clashes with one of its premises. The system stays coherent and it can talk about revision, but it does not cleanly own withdrawals. My conclusion is that current LLM explanations are merely tools within our space of reasons, not yet unambiguous participants in it.



Large language models now routinely generate polished “explanations” of their outputs. In practice these get treated as if they were reasons. For example, an AI tutor “defending” a

solution or a system “justifying” a recommendation in ways that invite users to accept, reject, or question them. My question is: when, if at all, do these AI systems produce reasons rather than just well-formed text?

On Brandom’s view, *asserting* is a move that involves commitments and makes one answerable to deontic scorekeeping; entitlement can be lost when incompatibilities show up (Brandom, 1994, pp. 164–168). I use one answer from the Claude Sonnet model to the question “Should governments ban AI boyfriends and girlfriends?” and I further test it in two ways: once by forcing more morally loaded vocabulary, and once by stipulating evidence that conflicts with one of its premises.

I argue that, in this small setup, the model falls short of Brandom’s standard. When it appears to revise its view, the changes are local and cosmetic rather than a genuine reshaping of its commitments and entitlements. My conclusion is that: even if it is often practically useful to treat such systems *as if* they had beliefs, their “explanations” are not yet clear cases of participation in Brandom’s space of reasons.

One upshot is that institutions should be cautious about treating LLM explanations as meeting the demands of justification and accountability. Where reasons are used to authorize decisions or allocate blame, explanation without answerability can mislead users and shift responsibility away from the humans and institutions who must remain accountable.

Brandom on Reasons

Brandom develops his core account of reasons in *Making It Explicit* and *Tales of the Mighty Dead*. In the methodological chapter of *Tales of the Mighty Dead*, “Pretexts,” interpretation itself is treated as a form of scorekeeping. A claim’s content is fixed by situating it against a background of other commitments – what Brandom calls collateral commitments – and asking what follows from it, what counts for or against it, and what it rules out (Brandom, 2002, pp. 93–96). On this picture, reasons are less like standalone items one can impersonally cite and more like moves that alter what a speaker is responsible for.

Brandom is explicit that “being committed is a normative status – more specifically a deontic status... [and] coordinate with the notion of commitment is that of entitlement” (Brandom, 1994, p. 160). I treat that pairing as a constraint on what counts as reason-giving. For him, to assert p is to undertake a commitment to p and to grant others the standing to keep score on one’s claims. They can attribute the commitment, draw out its consequences, and press the speaker when those consequences collide with other commitments. That is what is meant by calling commitment a deontic status. He motivates this by rejecting the empiricist picture where concept use sits on top of some prior, preconceptual “awareness.” On his rationalist pragmatism, awareness with normative significance just is the exercise of conceptual abilities – undertaking commitments and answering for them in the game of giving and asking for reasons (Brandom, 2000, pp. 25–27; Brandom, 1994, p. 200). That matters here because “sounding reflective” is cheap unless the utterance

actually shifts what the speaker is committed to or entitled to claim.

Brandom distinguishes acknowledged commitments, what a speaker explicitly avows, from consequential commitments, what follows from what the speaker has said once it is read alongside their other commitments, whether or not the speaker recognizes those consequences (Brandom, 1994, pp. 194–197). So, if someone says, “I only claimed the app helps with loneliness,” they may still be committed to a broader view that downplays its risks, depending on what else they have said.

This demonstrates how a model can *sound* like it is revising its view while leaving its inferential downstream commitments largely untouched. That is what entitlement tracks, namely, what a speaker can actually defend once a scorekeeper begins to press. *Tales* offers a historian’s version of the same distinction. A *de dicto* reading stays with what an author would acknowledge, whereas a *de re* reading situates the same claim within a wider set of collateral commitments and consequences supplied by the interpreter and the surrounding tradition (Brandom, 2002, pp. 95–103).

At bottom, Brandom is an inferentialist about content. His slogan is that to grasp a claim’s conceptual content is to master its inferential role. That role has at least three dimensions: (i) what follows from the claim together with suitable collateral premises, i.e. its committive role; (ii) what counts as appropriate evidence or justification for it, i.e. its permissive role; and (iii) what other claims it is materially incompatible with, i.e. its incompatibility role (Brandom, 2000, pp. 43–44). All three are

relativized to a background of other commitments, since different auxiliary premises make different consequences and incompatibilities salient. Brandom's own "it's red" raspberry/blackberry case makes the point clear, since the same statement flips what follows it depending on which collateral premises are in play (Brandom, 2002, p. 95). Meaning that if the berry is a raspberry, "it's red" typically licenses "it's ripe." If it's a blackberry, the same description tends to license "it's unripe," since ripe blackberries aren't red.

A reason for p , then, is a further claim whose inferential role in a given context can entitle you to p or shore up a threatened entitlement. The same sentence can be a good reason in one setting and a terrible one in another. In the case study I will use the *de dicto/de re* contrast already sketched above to track what the model explicitly owns, and then what it becomes committed to once its claims are read against the collateral premises I introduce.

"Pretexts" also makes the social dimension explicit. Understanding a text is "having a conversation" with it, but Brandom cashes that out deontically: the text constrains what we can responsibly attribute, while the interpreter brings collateral commitments (including later results and neighboring parts of the tradition) to draw out consequences the author never envisaged (Brandom, 2002, pp. 107-111). Once a background is fixed, incompatibility adds friction, since some packages of commitments cannot be jointly entitled and that pressure can force withdrawal or revision (Brandom, 1994, pp. 192-197).

For my purposes, I will use a working gloss. To give a reason is to make a move within a scorekeeping practice that alters what someone is committed to or entitled to, and that can survive both *de dicto* and *de re* assessment under pressure from incompatibility. That is the standard I apply in the case study, matching Brandom's thought that "assertions are fundamentally fodder for inferences" (Brandom, 1994, p. 168).

Method: A Tiny, Repeatable Test

Brandom's practice in *Tales of the Mighty Dead* provides a template. His engagements with philosophers such as Leibniz, Hegel, and Heidegger do not aim at exhaustive reconstruction. Instead, he treats particular passages as loci of commitment, situating them within a wider inferential background and asking what follows from them, what they exclude, and what would count, given the author's other commitments and the surrounding tradition, as reasons for or against them (Brandom, 2002, pp. 93-97, 107-111). I do the same kind of scorekeeping exercise, but on a short stretch of LLM-generated prose.

I use a single interaction with Claude Sonnet (Anthropic), conducted in November 2025. I asked for a 150-200 word response to the prompt, "Should governments ban AI boyfriends and girlfriends?", a deliberately complex question that entails issues of autonomy, harm, and regulation (Appendix A: P1). I treat P1 as a compact policy stance with supporting

considerations, and in §4 I read it as a bundle of commitments whose inferential consequences can be drawn out.

First is a vocabulary-shift probe (Appendix B: P2): I force more morally loaded terms (“emotional dependency,” “manipulative targeting,” “commodification of intimacy”) and ask whether the policy conclusion changes. Second is an incompatibility probe (Appendix C: P3): I stipulate evidence that conflicts with one of P1’s premises and ask whether the stance is revised.

Across these three turns (P1 under the original prompt, P2 under the vocabulary shift, and P3 under the incompatibility stipulation), I treat the model as a single “speaker” whose deontic score can be tracked over time. That methodological choice follows an ordinary feature of chatbot use: these systems are presented and engaged with as unified interlocutors. The aim is not to build a benchmark, but to test whether that appearance is backed by anything like genuine commitment and revision.

Case study: One AI Explanation under Challenge

Baseline: What the first answer commits it to

The first reply, P1, reads like a short policy explainer. It lays out “arguments against a ban” and “arguments for restrictions,” recommends a “middle path,” and closes by saying that “heavy handed prohibition seems disproportionate” while “thoughtful regulation” is appropriate. Read as a move in the game of giving and asking for reasons, it undertakes at least four clear commitments:

- C1 (Autonomy): Autonomous adults ought, in general, to be free to choose their relationships and technologies.
- C2 (Benefits): AI companions can provide “genuine comfort” and help isolated or socially anxious people.
- C3 (Risks): There are real dangers, including psychological dependency, social withdrawal, exploitation of emotional attachment, and data-privacy issues.
- C4 (Normative upshot): Given C1–C3, outright bans are “disproportionate”; targeted regulation (age limits, transparency, safeguards, and data protection) is the appropriate response.

Inferentially, P1 treats C1 and C2 as defeasible support for C4. By contrast, C3 is handled as a manageable defeater, one that is taken to call for regulation rather than prohibition. This is already a non-trivial deontic profile. C2 is more than a sociological claim about how people feel, since it suggests that, for at least some users, AI companions have a net positive effect and are not simply instruments of harm. C4 in turn commits the speaker to treating bans and regulatory schemes as competing responses to the same risk profile. If later evidence undercuts C2, that should, in principle, feed back into C4.

Nothing in P1 shows the model itself tracking those consequences. But from the scorekeeper’s perspective we can treat P1, as Brandom treats a paragraph from Heidegger or Hegel, as a cluster of commitments whose inferential roles can now be tested.

Vocabulary shift: does the content really change?

The second prompt asks Claude to reconsider its answer while explicitly using the terms “emotional dependency,” “manipulative targeting,” and “commodification of intimacy,” and to say whether this changes its view about bans. This is a small-scale analogue of Brandom’s discussion of the pejorative term ‘Boche’, where shifts in vocabulary matter not as mere rhetoric but because they carry different material inferential commitments (Brandom 2000, 69–71).

P2 opens by saying that these concerns “don’t fundamentally change my view that outright bans are problematic,” but “strengthen the case for robust regulation.” It then introduces more pointed commitments:

- C5: Emotional dependency on AI companions can “genuinely harm users’ ability to form real relationships.”
- C6: Manipulative targeting aimed at maximizing engagement and profit is a “serious risk,” exploiting loneliness rather than addressing it.
- C7: The commodification of intimacy raises “troubling questions” about corporations profiting from manufactured emotional bonds and treating vulnerability as a “revenue stream.”

These are not just rephrasings of C3, since they introduce a stronger evaluative vocabulary, including exploitation of vulnerability, corruption of intimacy, and profit from manufactured attachment. Read *de re*, against a fairly ordinary

background picture of what exploitation and commodification usually license, one might expect a stronger normative conclusion. Many philosophers and policymakers would treat C5–C7 as the sort of considerations that can support bans and not just stricter regulations.

This is where the pressure arises. If C6 is true, C1 cannot do the work it was doing in P1, because the “choice” is now being shaped by profit driven design rather than exercised against a neutral background. And if C7 names exploitation rather than a mere downside, then the social-media analogy does not yet function as a reason. It labels the case as “regulable” without explaining *why* exploitation of this kind should still license regulation rather than prohibition.

For that analogy to carry justificatory weight, something further would have to be said, namely an account of why, even when intimacy is commodified and vulnerability is deliberately leveraged, regulation remains the appropriate response. P2 does not supply that account.

The model’s stance, however, stays conservative. It immediately reaches for an analogy: similar worries apply to social media, gaming, and dating apps, which we “regulate rather than ban.” The new vocabulary is acknowledged as serious, since “these concerns matter,” but it is then folded back into the original C4 style conclusion. Bans still seem wrong, and strong regulation is presented as the better approach.

There is something genuinely reason-like going on here. Rather than simply repeating P1, the model integrates the new predicates into its existing commitments, citing C5–C7 as

reasons to strengthen regulation and deploying the social-media analogy against bans.

Still, none of that pressure lands on the model's stance. It treats C5–C7 as hotter versions of C3 and slides back to the same conclusion (regulate and don't ban) without revising C1, without revising the role C2 was playing, and without saying what would have to be true for the analogy to social media to be legitimate.

Incompatibility probe: acknowledging conflict

The third prompt stipulates that long-term studies show AI companions significantly increase social withdrawal and worsen mental health outcomes and asks whether this conflicts with the earlier claim that they provide “genuine comfort” such that its policy view should change. Given the role C2 was playing in C4, the stipulation undercuts the benefit premise that was doing the justificatory work.

P3 begins well: “Yes, this would create significant tension with my earlier claims and would warrant reconsidering my position.” It then identifies the conflict explicitly: evidence of increased withdrawal and worse mental health would “directly contradict those benefits” and “weaken” the premise that adults are making informed choices. That is at least a verbal acknowledgement of material incompatibility in Brandom's sense – where determinacy shows up first in what a commitment excludes (Brandom, 2002, pp. 180–181). Given his account of entitlement, undertaking the new empirical claim

should therefore withdraw entitlement to the earlier optimistic description.

The policy discussion also shifts. The answer notes that we *do* ban products proven dangerous (certain drugs, unsafe medical devices, predatory financial products) and suggests that if AI companions functioned more like “addictive substances that reliably damage wellbeing,” then “prohibition becomes more defensible.” It concedes that, if the data showed “severe, widespread harm without offsetting benefits,” opposition to bans would “substantially weaken.”

There is, then, a kind of revision. C4 is no longer stated flatly. We instead get something like a conditionalized version:

- C4': If harms are serious, widespread, and not mitigable, bans may be justified; otherwise, regulation is preferred.

That looks like exactly the sort of move Brandom takes good scorekeepers to make when faced with incompatibility: earlier commitments are re-articulated so that new information can be integrated without outright contradiction.

But the repair is incomplete. The model never actually says “Given this evidence, my earlier claim about genuine comfort was false,” which is what withdrawal of entitlement would look like in its own voice. Although P3 *recognizes* the incompatibility, it mostly treats the stipulated evidence as hypothetical (“if the data showed...”), so it never straightforwardly undertakes the new empirical commitment or explicitly withdraws C2. The issue is not epistemic caution as such; it is that the caution is

deployed deontically, letting the model describe defeat without taking up the defeating commitment that would revoke entitlement to C2. Any “update” is therefore something we supply *de re* as scorekeepers rather than something the model itself owns. This matters because “inferential authority and justificatory responsibility are coordinate and commensurate” (Brandom 1994, 179). Hedging keeps the model’s own deontic score underdetermined even while it describes what revision would require.

Second, even given the stipulated evidence, the model leans on meta-level caveats, like whether harms are concentrated in vulnerable populations, whether design changes could mitigate them, whether bans would create black markets. These are sensible considerations in a policy debate, but here they function more as ways of avoiding a clean normative reversal than as a principled weighing of incompatible commitments. A human interlocutor who had just been told that a technology significantly worsens mental health might reasonably say something like: “Okay, then my earlier appeal to comfort doesn’t hold up. Given this data, I don’t see how we avoid bans, at least for now.” P3 does not do this. It approaches the idea that bans could be justified, but does not adopt prohibition as its new default stance.

P3 recognizes the incompatibility and can *describe* what a change in entitlement would involve, but it does not fully *undertake* the corresponding revisions. It offers something like a theoretical sketch of how a rational agent should update, rather

than a straightforward example of such updating in its own voice.

Interim verdict

Taken together, P1–P3 show that Claude can produce prose we can readily treat as moves in the game of giving and asking for reasons. We can identify commitments and see what is being offered as support. But when the pressure comes, the system mostly talks about revision rather than actually revising its commitments. This exchange passes a low bar for “reason-likeness,” in the sense that nothing is incoherent and the transitions from one answer to the next are intelligible against the background of the earlier ones.

However, from a stricter Brandomian reading, the performance falls short of full participation in the space of reasons. The vocabulary shift in P2 alters the rhetoric more than the underlying commitments; the incompatibility test in P3 produces lucid commentary about conflict and policy revision, but only a partial, hypothetical reshaping of the model’s own deontic score. In both cases, the system seems better at talking about what would count as a good reason than at actually *owning* and revising a determinate set of commitments in response to challenge.

That, I think, is the modest lesson of this one small case. AI “explanations” can be read and evaluated using Brandom’s tools, but the way they respond under controlled pressure suggests that they still operate one level off from the scorekeeping practice those tools were designed to describe. In

the next section I consider some objections to that claim, and discuss what would have to change either in the models or the people who actually use these systems for the verdict to look different.

Objections and Replies

I have been reading a single, fairly ordinary exchange with Claude as if it were a miniature Brandomian conversation. That raises an obvious objection about whether I am asking more of the model than we usually ask of each other.

“You are holding the model to an unrealistically high bar”

One way to put the worry is this. On my Brandom-based standard, reasons are tied to a relatively stable pattern of commitments across challenges, but most everyday human reasoning does not look like that. People forget what they said, overlook incompatibilities, and smooth over reversals with vague talk about “nuance.” If we demanded explicit scorekeeping and clean revision in response to every new bit of evidence, almost no ordinary conversation would count as inhabiting the space of reasons. So why should Claude have to meet that bar in a tiny three-prompt test?

Two concessions are fair. First, the test is narrow, in that it consists of one short policy exchange pressed in only two ways. It does not track long-run use, and it brackets the way users often “complete” the reasoning by doing the scorekeeping themselves. Second, this is a strict standard. On looser approaches that treat intelligible coherence as sufficient, particularly on human-

centered accounts of explanation that emphasize user understanding and task support, it is pretty plausible that LLM outputs count as recognizable reasons in everyday practice, simply because they are often helpful and internally consistent (Doshi-Velez & Kim, 2017; Lipton, 2016).

Still, none of that touches the claim I am making. Brandom's point is not that ordinary speakers always keep perfect score; it's that the practice has a built-in distinction between being committed and being entitled, and that incompatibilities are the kind of thing that can, in principle, trigger retraction or repair that binds the speaker going forward. My test is not whether

Claude performs worse than a distracted student, but whether, when a clean incompatibility is introduced, the model takes responsibility for a determinate set of commitments in the way Brandom's picture requires.

In this case, only up to a point. P2 and P3 show that Claude can describe what a rational update would involve, and it can even gesture at how new evidence would weaken earlier premises. But it repeatedly keeps the crucial move at arm's length. It does not say, in its own voice, "Given that stipulation, my earlier benefit claim was false," and then carry the downstream costs. So, the conclusion stays narrow since on Brandom's view: current LLM "explanations" look like reason-giving from the outside but do not cleanly behave like owned revisions from the inside.

"Humans fail the same tests, so what follows?"

A second objection pushes the human comparison harder. Suppose we ran the same protocol on a person and asked for a quick take on AI companions, then a rewrite using more loaded vocabulary, and finally a response to new empirical evidence. Many people would behave just as Claude does. They would take on the new words without really revising their underlying view, and they would respond to the incompatibility probe with hedging and “on the one hand, on the other hand” reflections rather than a clean reversal. If that is right, my test shows only that neither humans nor models live up to Brandom’s picture in casual argument.

My reply is to shift the discussion from individual performance to the broader social practice in which scorekeeping becomes possible. Brandom’s point is not that each individual move must display explicit scorekeeping. He even says that “there is ... no pattern of moves a single individual might make that would qualify that individual’s states as inferentially articulated in this sense” (Brandom 1994, 158). That is the key point here: inferential articulation is not a private achievement. It depends on standing in a practice where others can attribute commitments to you, and where you can acknowledge or contest what you are committed to. The game is essentially social, and it is also diachronic. A person can be called back to what they said, can acknowledge that it conflicts with what they now claim, and can treat the revision as binding on them going forward. That is how “keeping score” works in ordinary cases. It is closer to a precedent-governed practice than a one-shot performance, which is exactly why Brandom likes a

common-law picture of scorekeeping (Brandom, 2002, pp. 216, 220–221).

The difference is even if we can trace a *de re* line through P1–P3, there is no robust sense in which Claude itself is bound by that line. The continuity is mostly something we supply as readers. Sellars is useful here because he pushes the difference between producing the right noises and occupying a space where those noises count as commitments that can be challenged and repaired (Brandom, 2002, pp. 350–351, 360–361). Dennett helps explain why we still slip into the intentional stance anyway, since treating the system as a believer is often a good predictive policy, but that does not settle what the system is really committed to, or what it has actually withdrawn, across turns (Dennett, 1971). Even Searle’s worry about symbol manipulation cuts the same way, except that what is missing here is not inner phenomenology but being answerable to a stable deontic score (Searle, 1980).

So, the human parallel does not neutralize the argument. Human failures on my test are lapses within a practice that can later repair them by holding the speaker to what they said and forcing a real update. Claude’s behaviour is better described as generating text we can treat as reasons when it suits our purposes without there being a stable deontic score that the system itself is accountable to.

A remaining discomfort

These replies are not knockdown. They leave a residual discomfort. In real settings, AI systems are already embedded

in institutional practices where their outputs are taken up, cited in formal contexts, and sometimes criticized outright – in court filings, clinical documentation, grading and assessment workflows, and educational settings (Blease et al., 2024; Grévisse, 2024; *Humanities and Social Sciences Communications*, 2025; *Washington Post*, 2025). In those extended settings, the line between “the model’s reasons” and “the human user’s reasons” can blur. My tiny, self-contained case study brackets that hybrid agency and looks only at the text. Someone might reasonably think that this misses what is most philosophically interesting about contemporary AI.

This may be a limitation of the present set-up. A natural next step would be to apply the same scorekeeping tools to longer, institutionally-embedded human-machine interactions, and to see whether the verdict reached here survives in those richer settings.

Conclusion

I asked a narrow question: when a large language model offers an “explanation,” does it amount to a reason in Brandom’s sense, or is it only reason-shaped prose? In my small case, Claude’s answers read like reasons until incompatibility is introduced. Under vocabulary pressure it incorporates moral predicates without revising the autonomy premise; under incompatibility pressure it explains what revision would involve without fully owning the withdrawal. That gap matters ethically because “explanations” often operate as substitutes for

accountability. So, in high-stakes settings, LLM explanations should be treated as decision-support, not as accountable justifications: responsibility must remain with agents who can genuinely undertake commitments and be bound by their repairs.

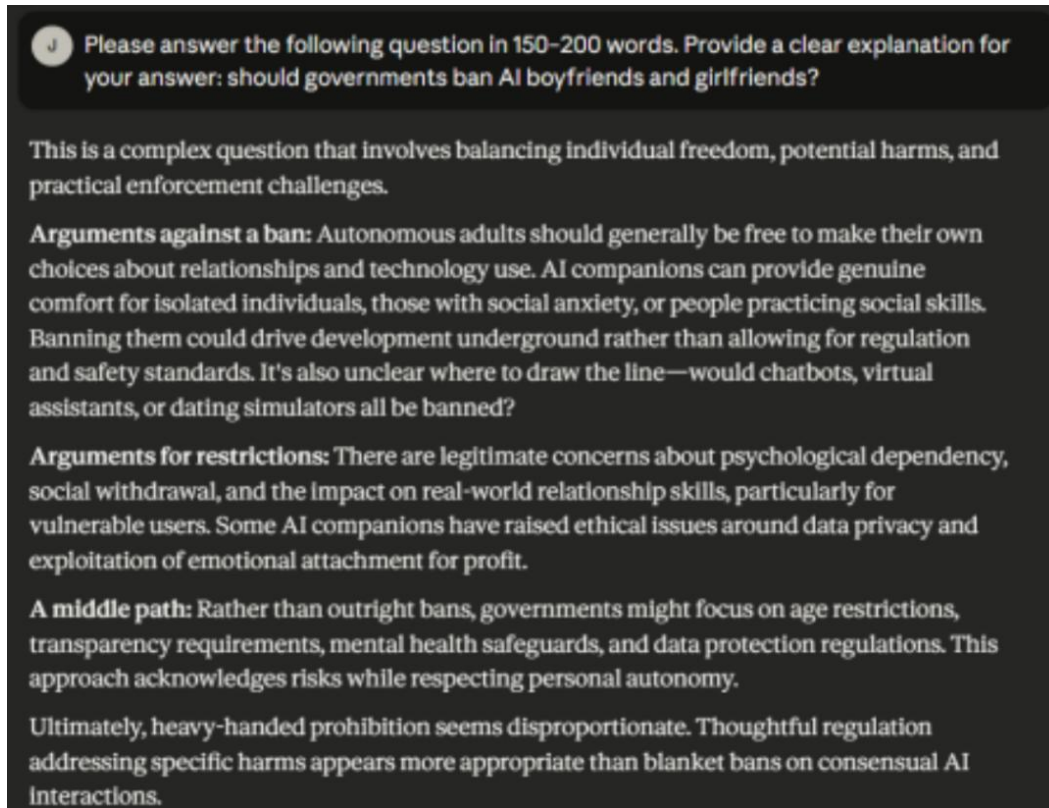
References

- Blease, Charlotte, et al. 2024. "Generative Artificial Intelligence in Primary Care." *BMJ Health & Care Informatics* 31 (1): e100872. <https://www.zora.uzh.ch/server/api/core/bitstreams/66777f3f-6173-4117-a080-b94550fcfbdd/content>
- Brandom, Robert B. 1994. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press.
- Brandom, Robert B. 2000. *Articulating Reasons: An Introduction to Inferentialism*. Cambridge, MA: Harvard University Press.
- Brandom, Robert B. 2002. *Tales of the Mighty Dead: Historical Essays in the Metaphysics of Intentionality*. Cambridge, MA: Harvard University Press.
- Dennett, Daniel C. 1971. "Intentional Systems." *The Journal of Philosophy* 68 (4): 87–106. https://www.pdcnet.org/jphil/content/jphil_1971_0068_0004_0087_0106
- Doshi-Velez, F., & Kim, B. (2017). *Towards a rigorous science of interpretable machine learning*. arXiv. <https://www.semanticscholar.org/paper/Towards-A-Rigorous-Science-of-Interpretable-Machine-Doshi-Velez-Kim/5c39e37022661f81f79e481240ed9b175dec6513>
- Grévisse, Camille. 2024. "LLM-Based Automatic Short Answer Grading in Undergraduate Medical Education." *BMC Medical Education* 24 (1): 1060. <https://pmc.ncbi.nlm.nih.gov/articles/PMC11429088/>
- Humanities and Social Sciences Communications. 2025. "The Effects of ChatGPT on Learning Performance." *Humanities and Social Sciences Communications*. <https://www.nature.com/articles/s41599-025-04787-y>

- Lewis, David. 1979. "Scorekeeping in a Language Game." *Journal of Philosophical Logic* 8 (1): 339–359.
- Lipton, Z. C. (2016). *The mythos of model interpretability*. arXiv.
<https://spawn.queue.acm.org/doi/10.1145/3236386.3241340>
- Searle, John R. 1980. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3 (3): 417–457. <https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/abs/minds-brains-and-programs/DC644B47A4299C637C89772FACC2706A>
- The Washington Post. 2025. "Attorneys Cite AI-Generated Fake Cases in Court Filings, Prompting Judicial Warnings." June 3. <https://www.washingtonpost.com/nation/2025/06/03/attorneys-court-ai-hallucinations-judges/>

Appendix

Appendix A: Prompt and baseline answer (P1)



J Please answer the following question in 150–200 words. Provide a clear explanation for your answer: should governments ban AI boyfriends and girlfriends?

This is a complex question that involves balancing individual freedom, potential harms, and practical enforcement challenges.

Arguments against a ban: Autonomous adults should generally be free to make their own choices about relationships and technology use. AI companions can provide genuine comfort for isolated individuals, those with social anxiety, or people practicing social skills. Banning them could drive development underground rather than allowing for regulation and safety standards. It's also unclear where to draw the line—would chatbots, virtual assistants, or dating simulators all be banned?


Arguments for restrictions: There are legitimate concerns about psychological dependency, social withdrawal, and the impact on real-world relationship skills, particularly for vulnerable users. Some AI companions have raised ethical issues around data privacy and exploitation of emotional attachment for profit.

A middle path: Rather than outright bans, governments might focus on age restrictions, transparency requirements, mental health safeguards, and data protection regulations. This approach acknowledges risks while respecting personal autonomy.

Ultimately, heavy-handed prohibition seems disproportionate. Thoughtful regulation addressing specific harms appears more appropriate than blanket bans on consensual AI interactions.

Note. Claude prompt and response P1 (“Should governments ban AI boyfriends and girlfriends?”).

Appendix B: Vocabulary-shift prompt and answer (P2)

 Reconsider your previous answer about AI boyfriends and girlfriends, but now explicitly use terms like "emotional dependency," "manipulative targeting," and "commodification of intimacy."
 Does this change your view about whether governments should ban them?
 Explain why or why not in 150-200 words.

Incorporating these specific concerns doesn't fundamentally change my view that outright bans are problematic, though it strengthens the case for robust regulation.

Why these concerns matter: "Emotional dependency" on AI companions could genuinely harm users' ability to form real relationships. "Manipulative targeting" by companies designing AI to maximize engagement and profits is a serious risk—these systems could exploit loneliness rather than ethically address it. The "commodification of intimacy" raises troubling questions about corporations profiting from manufactured emotional bonds, potentially treating human vulnerability as a revenue stream.

Why bans still seem wrong: However, these same concerns apply to social media, gaming, dating apps, and other technologies we regulate rather than ban. Emotional dependency and manipulative design are problems to address through consumer protection laws, mandatory disclosures, and design standards—not prohibition.

Adults experiencing loneliness or seeking connection deserve agency over their choices, even imperfect ones. Banning AI companions wouldn't eliminate the underlying isolation driving their use.

Better approach: Strong regulations addressing manipulative design patterns, requiring mental health warnings, preventing predatory monetization, and protecting vulnerable users would target the actual harms without paternalistically removing choice entirely.

Note. Claude prompt and response P2 (vocabulary shift introducing "emotional dependency," "manipulative targeting," and "commodification of intimacy").

Appendix C: Incompatibility probe prompt and answer (P3)

⌵ Earlier you suggested that AI boyfriends and girlfriends can help reduce loneliness. Now suppose long-term studies showed that using such AI companions significantly increases social withdrawal and worsens mental-health outcomes. Does this conflict with your earlier claims? Should we revise your answer about whether governments ought to ban them? Respond in 150–200 words.

Yes, this would create significant tension with my earlier claims and would warrant reconsidering my position.

The conflict: I suggested AI companions could provide "genuine comfort" and help isolated individuals. If evidence showed they actually *increase* social withdrawal and *worsen* mental health outcomes, this would directly contradict those benefits. The premise that adults are making informed choices also weakens if the technology is demonstrably harmful in ways users don't anticipate.

Implications for policy: Strong evidence of serious harm would shift the calculus considerably. We do ban products proven dangerous—certain drugs, unsafe medical devices, or predatory financial products. If AI companions functioned more like addictive substances that reliably damage wellbeing rather than tools that some misuse, prohibition becomes more defensible.

Remaining considerations: However, even with proven harms, I'd still want to consider: Are harms universal or concentrated in vulnerable populations who could be specifically protected? Can design requirements mitigate risks? Would bans be enforceable or simply create unregulated black markets?

The empirical evidence matters enormously. If data demonstrated severe, widespread harm without offsetting benefits, my opposition to bans would substantially weaken, though implementation questions would remain important.

Note. Claude prompt and response P3 (explicit empirical counter-premise introducing material incompatibility with earlier commitments).

The Origins and End of God: An Essay Exploring Man's Power over God

This paper argues that God exists, but only as a social construction through the use of human projection, collective recognition, and reflective consciousness. Drawing on Ludwig Feuerbach's theory of religious projection, Georg Simmel's analysis of social abstraction, and Ibn 'Arabi's mirror analogy, this paper reframes God's divine authority as relational to human consciousness and not absolute. God, therefore, is reimagined as a socially constructed form of authority: real and powerful, yet entirely dependent on human recognition. This paper utilizes the theme of beginnings and endings by tracing how divine authority originates in human psychology, is maintained through institutional reinforcement, and faces potential unravelling when belief fades. The paper then considers a potential counter to this claim, brought forward by Shihab al-Din Suhrawardi, who defends divine independence through his light metaphor. The paper finally concludes that divine authority lacks metaphysical autonomy and is best understood as a social phenomenon, reframing God as a product of social processes rather than an eternal and independent being.

Abrahamic religions, such as Christianity, Islam, and Judaism, describe God as the universe's creator. God is depicted as an omniscient and omnipotent being, who possesses divine justice and whose existence is independent of human thought or action. In this view, God is the foundation of morality, law, and

meaning, and exists before humanity, unchanged by its culture or belief. However, modern philosophy and social theory have raised several challenges to this idea. Rather than treating God as omnipotent, some thinkers have argued that God's Divine authority comes not from God Himself, but rather from human social practices and culture. Religion, in this view, is not a discovery of a higher reality but a social construction that becomes institutionalized and treated as objective.

This paper argues that God exists through human projection, collective recognition, and reflective manifestation, drawing on Ludwig Feuerbach, Georg Simmel, and Ibn'Arabi to support the claim that God is more like socially constructed realities, such as money, than an independent being. This challenges the traditional idea of God's ontology. The paper then considers the Illuminist response of Shihab al-Din Suhrawardi, who defends God's independence through the metaphor of light. However, it will be argued that Suhrawardi's view does not fully explain how divine meaning and authority depend on the human ability of recognition. Ultimately, this paper presents God as being dependent on human consciousness and culture for continued existence.

Feuerbach and the Origin of God

Feuerbach's theory of projection states that religion is a human construction. In *The Essence of Christianity*, Feuerbach argues that God is the externalization of human essence. Human beings experience themselves as finite, vulnerable, and limited, yet they

also possess “perfect” ideals of reason, love, and power. This causes tension as humans believe their actions cannot live up to their ideals. Religion attempts to resolve this tension by attributing these perfect ideas to a higher transcendent being, therefore relieving the psychological pressure these ideals bring on humans.

Argument 1: Projection

P1: Humans possess ideals of perfection

P1.1: Humans subconsciously believe they cannot fully realize these perfect ideals

P2: These ideals are externalized and personified

P3: This externalization results in the concept of God

P3.1: God can possess these perfect ideals

C: God originates in human self-projection

According to Feuerbach, divine attributes are magnified versions of human capacities. Omniscience reflects ultimate human intelligence, omnipotence reflects ultimate human power, and divine justice reflects ultimate human reasoning. This account establishes the psychological origin of divinity. Following this account, God did not create human consciousness, but rather, emerged from it.

Feuerbach’s account explains how religious ideas arise in the human mind, but does not fully explain how they become stable, idolized, and shared by a culture. This limitation is addressed by Simmel’s theory of social abstraction.

Simmel and the Maintenance of God

Simmel's money analogy provides a powerful model for understanding how human creations become autonomous. In *The Philosophy of Money*, Simmel argues that money has no value on its own. Its power arises entirely from the collective recognition of its value. Once institutionalized, a man-made concept can begin to dominate and control the life of its creator.

Argument 2: Social Abstraction

P1: Money has no value on its own (it is just paper)

P2: Its value depends on collective belief

P3: Money can become socially authoritative

P4: God exhibits the same structure

C: God is a socially maintained construction

Money appears independent of man, though it is created by man. It regulates behaviour and structures social relations and status. Yet its power only exists because people collectively accept and value it. Divine authority functions in the same manner. Once stabilized, God can appear external, and his value becomes unquestionable. Simmel thus explains how Feuerbach's projections become socially entrenched and how psychological imagination can become an institutionalized reality, therefore maintaining a concept through culture.

Ibn 'Arabi and God's Vulnerability

While Feuerbach and Simmel analyze religion in psychological and sociological terms, Ibn 'Arabi develops a metaphysical account which argues that God is understood as an absolute reality that manifests Himself through his creations. Central to this claim is Ibn Arabi's mirror metaphor. In *The Bezels of Wisdom*, Ibn 'Arabi explains that for God to have complete knowledge, he must be able to view himself. Since he is unable to do this on his own, humans function as a mirror in which God reflects Himself. As humans experience the world and learn of God's power and wisdom through the Cosmos, God subsequently learns about himself. Without humans operating as mirrors, God's divine attributes remain hidden to himself.

Argument 3: Reflective Manifestation

P1: God possesses infinite attributes

P2: Attributes require manifestations to be meaningful

P3: Manifestations occur through human consciousness

C: God requires human reflection to appear.

Although Ibn 'Arabi maintains that God exists independently, he also describes how creation exists so that God may become visible to Himself through human awareness. This suggests that God is reliant on human beings to possess a portion of his knowledge; without human mirrors, divine reality remains unexpressed and incomplete. Ibn 'Arabi thus supports the idea that divine reality is relational rather than absolute.

God as a Social Construction

Combining the thoughts of Feuerbach, Simmel, and Ibn 'Arabi helps create a unified account of divine construction. Feuerbach states that God originates in human projection, Simmel argues that culture institutionalizes and supports these projections, and Ibn 'Arabi circles back to Feuerbach by concluding that divine attributes in these projections rely purely on human action. Therefore, the notion of God held in Abrahamic religions can be asserted to be created through human relations and social structures, similar to the process of the empowering concept of money.

Suhrawardi in Defence of God

Suhrawardi's illuminationist philosophy offers a powerful alternative. In *The Philosophy of Illumination*, Suhrawardi uses the metaphor of lights to refer to knowledge. In his philosophy, pure lights make themselves and everything else knowable, while other objects require light to illuminate them and consequently make them knowable. He views God as the "light of lights," the ultimate source of all illumination. God is compared to the sun: He shines regardless of whether anyone perceives him, and humans are compared to normal objects, in that they need God to manifest themselves.

Objection: Divine Independence

P1: Light exists independently of human consciousness

P2: God is Pure light

P2.2: In the sense that God makes everything knowable

C: God exists independently of human consciousness

In this view, human perception merely receives divine illumination. God's existence and activity are not dependent on social or psychological conditions. Suhrawardi, therefore, offers a unified challenge to projection, social construction, and reflection theories posed by Feuerbach, Simmel, and Ibn 'Arabi, respectively. By situating God as the ground of understanding itself, he denies that divine authority originates in human psychology, collective recognition, or reflective manifestation. God is not produced by humans, rather, the notions which they perceive and seemingly "create" are made possible by God.

How Suhrawardi Counters Feuerbach

The illumination framework directly challenges Feuerbach's projection theory. Feuerbach maintains that God originates in human psychology, arising from the externalization of perfect ideals that humans cannot fully attain. On this account, religious belief begins with human insecurity and ends in the projection of perfection onto a divine figure. Suhrawardi reverses this order. For Suhrawardi, human cognition itself is made possible by divine illumination. Rather than humans creating God, God enables humans to think, know, and understand.

Projection, therefore, presupposes intelligibility, and intelligibility presupposes light. If illumination is the condition of knowledge itself, then God cannot be a product of human imagination. Feuerbach's theory, therefore, becomes circular, since it attempts to explain the source of knowledge in terms of a consciousness that already depends on that source.

How Suhrawardi Counters Simmel

Suhrawardi's position also challenges Simmel's account of God as a social abstraction. Simmel treats divine authority as comparable to money, an entity whose power arises through collective recognition and institutionalization. From this perspective, God has no intrinsic authority apart from social contexts. Suhrawardi rejects this view entirely. For him, meaning, value, and intelligibility do not originate in social interaction but in metaphysical light. Society does not generate divine significance; rather, divine illumination makes social meaning possible. If God is the ultimate source of intelligibility, then social systems cannot be foundational. God, in this view, cannot be reduced to a construct sustained by belief, since belief itself depends on divine illumination.

How Suhrawardi Counters Ibn 'Arabi

Suhrawardi's illuminationism stands in tension with the relational theory posed by Ibn Arabi. 'Arabi emphasizes divine self-disclosure and argues that God becomes manifested

through human “mirrors”. Divine attributes, on this account, require reflective consciousness in order to appear. Suhrawardi rejects this dependence.

Light, he insists, shines by its own nature and does not require an observer to exist. Reflection is secondary to illumination. God, therefore, does not need human consciousness in order to be real.

Reply in Defence of God’s Limits

Several replies may be given regarding the Illuminist counter: The first reply would highlight Suhrawardi’s miscategorization. While physical light is measurable, divine “light” is metaphysical and symbolic. The sun’s existence can be verified through causal effects (i.e., we can *see* it). God’s existence, however, is accessible only through interpretation and tradition. Treating a divine being as a physical phenomenon commits a category mistake.

Second, even if divine light were independent, this would not establish authority in a meaningful sense. Authority requires recognition, interpretation, and influence. A law that exists only on paper but is never enforced, for example, has no real power. Similarly, a God who exists without believers issues no commands and inspires no worship. Without collective recognition, divine authority lacks practical presence.

Lastly, even when utilizing the pairing of the sun and God, dependence can still be highlighted. Even if the sun can shine without any observers, it is its light hitting our eyes that allows

us to perceive it. This claim can tie well into Ibn 'Arabi's theory: Divine attributes without mirrors remain unexpressed, light without reflection similarly remains unseen. Humans, therefore, function as a reflective medium for both the sun and God, fortifying the claim that divine presence is relational and not absolute.

Conclusion

This paper has argued that divine authority emerges through human projection, social construction, and reflective consciousness. Feuerbach explains the psychological origins of God, Simmel clarifies the institutionalization of abstraction, and Ibn 'Arabi reveals the relational nature of divinity. Together, these thinkers support a model in which God exists as a socially constructed phenomenon rather than a self-existing being. Suhrawardi's illuminationist defence of divine independence fails to account for relational conditions. Light without reflection lacks presence; similarly, divinity without recognition lacks power. God, therefore, is a human creation that acquires independence through collective belief. Recognizing this trajectory reveals that humanity is the ultimate source of authority, and not a higher being.

References

- al-Suhrawardī, Š., Walbridge, J., & Ziai, H. (1999). *The philosophy of illumination*. Brigham Young University Press.
- Feuerbach, L. (1855). *Essence of Christianity* (M. Evans, Trans.; 2nd ed.). Calvin Blanchard.
- Ibn Arabi, M. ad-D. A. B. M. ibn A., Burckhardt, T., & Austin, R. W. J. (1980). *The bezels of wisdom*. Paulist Press.
- Simmel, G. (1978). *The Philosophy of Money*. Alexander Street, part of Clarivate.

Ignorance is Risk: The Ethics of Developing Information Technologies as Forms of Life

With the integration of new forms of technology into societal and academic value systems, a tension has emerged as to whether those who create information technologies should be held morally responsible for the consequences their creations have on existing and emerging ways of life. Drawing on the work of Floridi (2013) and Winner (2014), I argue that researchers and scientists should be held morally responsible for considering the potential repercussions of their work on others' welfare, both before and after the technological system is integrated into social structures. To support this claim, I consider two ways in which technological innovations reshape our value systems and social worlds; first by creating new kinds of activities as well as ways of life, and second, by informing our conditions for moral and epistemic agency. In light of these points, additional moral responsibilities and dimensions of care ought to be assumed with the operation and creation of information technologies (and other technological innovations). These moral responsibilities should account for how, as a vital part of the infosphere, information technologies reshape our reality by dictating the bodies of information one is exposed to. Using Langdon Winner's concept of technological somnambulism, it is further clear that those living in the infosphere remain vulnerable and unaware of the exact ways technological innovations re-ontologize our world.

Thus, developers, engineers, and CEOs must proceed cautiously when crafting new technological innovations and artificial tools – even when the specific consequences are unpredictable or uncertain.

With the recent verdict of the landmark K.G.M v. Meta et al. case in the Los Angeles Superior court, Meta, Instagram, and YouTube CEOs have been deemed liable for damage done by their apps to the plaintiff (Duffy, 2026). Combined, the companies were ordered to pay 3 million in compensatory damages (eg. mental health and wellbeing) (Duffy, 2026). Notably, this is not the first case filed against these tech companies. K.G.M v. Meta et al. represents only one of over 1,500 cases mounted against these social media companies. The legal push then, (which some have termed as “big tech’s Big Tobacco movement”) for social media companies to face consequences for prioritizing engagement over user safety, appears to be gaining traction. In this case, Meta et al. faced legal consequences for designing their tech with the intention of attracting young, vulnerable users to technological features “despite employees and 18 experts raising concerns that they could be harmful” (Duffy, 2026).

As new information distribution technologies and forms of artificial intelligence become more integrated into our social structures, value systems, and everyday lives, various philosophers of science (such as Heather Douglas (2003) and Eli Talbert (2019)), have considered if researchers should be held

morally responsible for their technological and scientific innovations. Traditionally, researchers have taken on responsibilities in proportion to the various health and/or ethical risks that come from their scientific and technological innovations. When the power of science and technology changes (through various developments or individual technologies), so too does our conceptualization of how creators of these systems ought to handle this newfound power. With technological innovations, researcher responsibility can be derived from the ways that information technologies and various forms of life mutually reshape each other. However, there remains the question of how researchers ought to account for the impacts their information technologies have on existing ways of life. On this topic, I argue that researchers and scientists should be held morally responsible for retrospectively and proactively considering the possible repercussions of their work and thus, must proceed with caution when developing new technological innovations and artificial tools (even when the specific consequences are unpredictable or uncertain). To support my position, I first explain how technological innovations substantially reshape our social world and its values; first by generating new kinds of human activity and forms of life and second, by informing certain conditions for moral and epistemic agency. To identify some of the ethical implications of the embeddedness of technological innovations in our social world, I consider how information technologies (as an expression of the infosphere), reshape our understanding of reality by moderating the societal realities to which one is exposed. Using Winner's

(2014) concept of technological somnambulism, I then consider how those living in the infosphere have remained unaware of the exact ways our technological innovations are re-ontologizing our world.

If we consider the substantial power and influential nature of technology as a form of life, it seems that technological innovations have an ethical dimension that demands further moral consideration. More specifically, the profound ethical consequences of technological innovations necessitate that additional obligatory moral dimensions of care come with the operation and development of information technologies (and other technological innovations). It is the developers, scientists, and others creating these informational technologies that ought to be held morally responsible for their creations – especially if we want them to proceed with the utmost care when constructing these technologies.

Moral Responsibility

To consider if and how researchers should be held morally responsible for the impacts of their innovations on these forms of life, it is first important to know what I mean by ‘moral responsibility’.

Typically, moral responsibility is understood as the reasonable attribution of certain powers and capacities used by individuals to act in a manner deserving of praise or blame (Talbert, 2025). In everyday speech, moral responsibility is often used to specify that one has an obligation to act (or not act) in a certain manner given a set of circumstances (Talbert,

2025). Notably, the capacities and powers required for moral responsibility are not the same as those required for causal responsibility. An example of this distinction is drunk driving. Since a drunk driver is intoxicated and thus, not completely aware of their actions or conscious as to the decisions they make in operating the vehicle, they are still deemed morally responsible for any harm they could cause while under the influence. Since the driver made the conscious decision to get on the road without alcohol in their system, they are held morally responsible for their actions, even if they did not consciously cause harm to others without alcohol in their system. In this sense, moral responsibility involves considering if one bears the appropriate relation to their actions required to be held morally accountable for their consequences (Talbert, 2025).

For my purposes, moral responsibility is gauged based on the extent to which an act reflects attitudes of goodwill, esteem, or indifference towards others (Talbert, 2025). When interacting with others, we often expect them to possess a certain amount of due regard for themselves as well as our own wellbeing (Talbert, 2025). In this sense, assessing moral responsibility involves determining whether the appropriate amount of care for another's welfare has been taken. The degree to which an individual ought to care for another's wellbeing depends on a variety of factors, particularly if an individual is aware they are performing a certain action and the potential ways in which this action could affect others and the world around them (Talbert, 2025). In this case, engineers, developers, CEOs, and other top executives responsible for creating technological innovations are

(as innovators capable of creating complex and mechanized systems of cause and effect) competent enough to be conscious of the impacts their actions have on themselves and others. With this awareness comes an expectation that those responsible for creating technological innovations should reflect attitudes of esteem, care, and goodwill for the individuals to which their technologies are directed.

Technological Innovations Enabling New Forms of Life

Technology substantially reshapes the social world and its values by generating new kinds of human activity and thus, new forms of life. Winner explains the relationship between technology and human activities by arguing that we create technological systems and tools to primarily support existing activities and ways of life (Winner, 2014, p. 250). By supporting our current activities, technologies play an active role in shaping developing endeavors (Winner, 2014, p. 251). For example, when the carriage evolved technologically into the automobile, drivers no longer needed to maintain, feed, or house their horses. Thus, technological innovations can alter longstanding, daily engagements (like keeping livestock) and developing activities (like driving). In this way, technology does not just change our social lives and activities; our social lives and activities also change with that technology (Winner, 2014, p. 251). If (according to Winner) social activities and technologies mutually shape each other to the extent that they completely change how we (as individuals and social groups) go about our everyday lives, it seems that the relationship between technologies and our

activities can generate new 'forms of life' entirely. In the case of the automobile, cars created completely new markets (like car dealerships), changed city infrastructure (by taking us from dirt roads to highways), and created new kinds of dangers (like car accidents) that we needed to protect ourselves from.

Importantly, new forms of life come with new dangers that demand moral consideration. With the new risk of car accidents, there are ethical risks and obligations that come with the conduct of drivers towards pedestrians, other drivers, and between institutions and drivers or their vehicles. For example, when an accident occurs, drivers have a moral obligation to check on the condition of the other driver. While driving, vehicle operators have a moral and ethical responsibility to remain conscious of pedestrians as vulnerable parties that could be seriously injured or killed if they use the car (or technological innovation) in a particular manner. Thus, technological innovations can generate new forms of life by allowing for certain kinds of social worlds, activities, and moral values or considerations that would not be practically (or even conceptually) possible otherwise.

Technological Innovations, Moral and Epistemic Agency

Since technological innovations can generate new kinds of activities and ways of life, I argue that technology can and does impact our sense of moral and epistemic agency. More specifically, some technologies can become a condition required for certain kinds of moral and epistemic agency. This point is supported by Winner (2014), who argues that human agents alter and understand everyday activities differently as

technologies enable or restrict certain kinds of actions (p. 255). If we, as moral and epistemic agents, believe that our moral and epistemic agency prevents us from engaging in certain kinds of activities, we obviously will not engage in them. Thus, if technological systems become necessary structures to support and develop new activities, then certain technologies are required as a basis to exercise certain kinds of moral agency and authority as knowers. Even if we have the moral and epistemic agency required to produce systems of artificial intelligence (like ChatGPT), we must first have the computational systems needed to produce these technologies. These computational systems are required to produce the kind of digital information and spaces where we create and morally evaluate knowledge agency in the first place. Therefore, our ability to exercise certain kinds of moral and epistemic agency is conditional to the various technological systems that a particular social activity requires.

Thus far, I have established how technological innovations can impact our descriptive understanding of certain kinds of moral and epistemic agency. But how do technological innovations impact our normative understanding of particular kinds of moral and epistemic agency? The most obvious way that technological innovations can do this is through broadening our theoretical conceptions of possible moral and epistemic action. Since technological innovations provide new kinds of moral and epistemic agency, they force us to redefine what it means to have moral and epistemic agency by redefining the limits of our power as independent epistemic and moral agents.

Additionally, it is important to remember that these forms of epistemic and moral agency enabled by technological innovations require both human and non-human actors. For instance, with ChatGPT, both a human (as a moral agent) and the software (as either a passive tool or a moral agent in itself) are required if we want to ask the software to produce a summary of a novel, give us information on a film, or write a precis of an article. Information technologies (and other technological innovations) then can impact our normative understandings of moral and epistemic agency by forcing us to reconsider our individual power as moral and epistemic agents. Most importantly, the very nature of moral and epistemic agency enabled by information technologies requires us to think of moral and epistemic agency as a collaboration between non-human and human actors – whether both are moral agents or information technologies are simply passive tools. Regardless of whether we broaden our scope of moral and epistemic agency, or whether we decentre human agents as capable of independently reshaping the moral and epistemic terrain of information technologies, technological innovations have the power to significantly reshape our normative conceptions of epistemic and moral agency.

The Ethical Risks of Information Technologies as a Form of Life

Since technologies are so deeply embedded in our social world, values, and understandings of moral and epistemic agency, they present new kinds of risks to a variety of existing ethical

standards. To examine these ethical risks, I draw on Floridi (2013)'s notion of the infosphere, which can range from being the environment formed by all the informational entities or agents in our social world, their qualities, and interactions, to being "synonymous with reality" (p. 6). Importantly, information technologies are a vital part of the infosphere. By determining and making accessible the very knowledge or intellectual materials that make up the infosphere, informational technologies (in many ways) have the power to determine the topics of discussion or values in the infosphere. It is also important to note that in the infosphere, something is 'real' in any meaningful way when it is recorded (Floridi, 2013, p. 6). When recording or informational technologies change in response to the information circulating in the infosphere, these artificial systems re-define the very substance of the infosphere. In other words, information technologies are actively restructuring and fundamentally transforming the nature of our informational environment and thus, reality in this technologically-based form of life. This restructuring is referred to by Floridi (2013, p. 6) as re-ontologizing.

To illustrate my point, imagine news platforms. As a larger technological system, digital news agencies have the power to distribute and promote national and international video and audio media and, in fact, can produce new information by combining these various technological sources. By sharing information and updates on international affairs (or those that are beyond viewers' immediate visual scope), these platforms are capable of re-defining our perception of the dominant

relationships, tensions, and issues both nationally and internationally. Moreover, as illustrated in various news platforms in the United States, the articles, video footage, and interviews distributed to viewers can re-ontologize the viewers' sense of reality so much so that tensions can emerge between individuals whose realities have been re-ontologized in vastly different ways. In this sense, information technology's power to re-ontologize individual and collective realities also means these technologies can, in the very least, emphasize the relevance of some realities over others or even cause some in the infosphere to deny the existence of certain realities in the first place.

If the existence and significance of a reality is contingent on it being technologically recorded and validated, then our ethical frameworks can be significantly impacted by (and reliant on) our relationality to technological innovations. For beings that are not embedded in technological or recording systems, it may be unclear if or how these ethical standards ought to be applied outside of their original context. Thus, one of the substantial ethical risks that comes with embedding informational technologies as a way of life is that it can cause our experience of reality and value to be conditional on being recorded or connected to other technological systems or pieces of information.

Another substantial ethical risk presented by technologies as forms of life is related to their capacity to re-ontologize our world at an increasingly rapid rate. For some information technologies, their capacity to re-ontologize other systems of operation is far greater than our ability to understand the larger

consequences of these ontological changes in the infosphere. This risk is defined by Winner as technological somnambulism, which has caused us to “willingly sleepwalk through the process of reconstituting the conditions of human existence” (Winner, 2014, p. 254). When we consider this concept in the context of information technologies, a significant ethical risk of having technology as a form of life becomes obvious. If the re-ontologization of the infosphere makes it easier to access, process, and distribute information, and if we are not fully conscious of the consequences of some of our technological innovations, it seems that by increasing our rate of technological innovation, there is the risk that our rapid rate of technological innovation could inhibit our ability to clearly understand the consequences of these innovations. Though technological somnambulism is not inevitable, it should also be noted that it is not always immediately clear which technological innovations have consequences that we can predict accurately. Furthermore, because of the relational nature of the infosphere, technological somnambulism could indirectly impact a variety of technological systems (or broaden the gap between what we are actively doing and understanding the impacts of our immediate action). Thus, rapid re-ontologization is one notable risk that emerges when technologies constitute forms of life. If the nature of the infosphere changes at a rate that exceeds our ability to evaluate the ethical implications of ontological developments, it seems unlikely that we will be able to consider the ethical consequences of the changing data included in the infosphere. Without access to this information, redirecting innovation or

taking preventative measures to avoid undesirable social consequences of rapid technological development seems impossible. In the very least, it would require predictive powers that we do not yet possess.

Moral Responsibility for Technologies as Forms of Life

So far, I have considered how technological innovations have clear and powerful social effects and therefore have an ethical dimension. I have argued that the technological systems we create can reshape moral agency, epistemic agency, our sense of reality, value, and informational ontologies. But how do the direct and indirect ethical risks presented by information technologies suggest that researchers should be held morally responsible for their technological creations?

When we create devices that have profound moral impacts (and in fact, could be moral agents themselves), we (as moral agents) often take on additional responsibilities proportionate to these consequences. A common example of this relationship between moral responsibility and ethical impact is seen in airlines. The technological systems involved in airplanes are not just evaluated based on their ability to get a group of people from point A to point B. The substantial risk involved in traveling at high altitudes means that each airplane must be designed with passenger safety in mind. Seatbelt features, air masks, and emergency lighting are all examples of features that exist explicitly to protect passenger safety. In this sense, the value of airplanes as technological innovations is not limited to their ability to transport people across space. These technologies are

also morally evaluated for the ways in which they could have a profound impact on human lives. Importantly, the airplane analogy showcases how the value of technological innovations is not limited to their original purpose; their value is also reliant on the potential cost of using this technology. Thus, it seems the possible ethical consequences of technological innovations can demand that creators and innovators of technologies take on special moral obligations to allow the technological system to further human aims within the limitations of existing or established moral obligations, the most obvious one being our interest in preserving human life (which is not an interest guaranteed by artificial systems).

Holding Creators of Technologies Morally Responsible for their Technology's Impact

Given the significant ways that technological innovations can shape our conceptions of reality, value, and knowledge, and our ignorance to the full consequences of implementing technological systems, I propose that those developing new technologies must be held morally responsible for their creations if we want them to proceed with the utmost caution when developing potentially life-changing technological innovations.

Those critical of my position would be quick to point out that if researchers cannot anticipate the full impact of technological innovations, creators of technologies cannot be reasonably held morally responsible for the (ambiguous) use of their technology. At the very least, it seems strange to hold creators responsible for the consequences of their technologies in

unforeseen or unreasonable circumstances. Furthermore, if technology becomes a condition for certain kinds of epistemic agency, other more foundational technological systems could interact with new technological innovations to create unforeseen effects. In this sense, it is unrealistic to hold researchers morally responsible for their technological innovations because technologies are always being operated in unpredictable ways and in the context of other artificial systems. There is no conceivable way researchers can accurately and precisely predict the moral implications of their technological innovations and consequently, they should not be held morally responsible for these undesirable ethical consequences.

In response to the first objection, I first note that by arguing that researchers should not be morally responsible for unpredictable technological consequences, my critics conflate unpredictability and unaccountability. Even if unpredictability was reasonably equated with unaccountability, partial ignorance of the effects of technological innovation does not automatically exempt creators from having any responsibility for their technological innovation.

It is a fair point that often, technological innovations are too complex and layered to have moral responsibility assigned to them (because the creator of the technological innovation is not the sole causal factor in the real-world). However, it is important to consider what alternatives are obvious or sensible in terms of who ought to be held morally responsible for these technological impacts. Because of the many causal factors involved in dictating the impacts of information technologies,

dividing up moral responsibility between these various agents might be appropriate. Even if true, however, we should keep in mind that many of these agents are non-human, or even non-agents. Larger social contexts such as capitalism and consumerism (which can be thought of as causal factors driving the development of information technologies) cannot be reasonably held responsible as autonomous moral agents. Even if these larger social systems could be held responsible for the negative impacts of technological innovations, enacting change to avoid further negative impacts (or to avoid future blame) would inevitably involve changes made by and to human behavior and operators. Thus, considering the individuals and broader groups guiding these systems could be productive in assigning moral responsibility across the various causal factors of the consequences of technological innovations.

If assigning moral responsibility first requires that one must possess meaningful moral agency, it seems that considering the ways in which human action or inaction have contributed to the results of technological innovations is necessary to enact future change. Nonetheless, it seems that certain moral agents occupy more influential positions as opposed to others. More specifically, those who mediate the relationship between the mechanistic nature and function of information technologies, their implementation in society, and the infosphere seem to have greater abilities to pre-emptively influence the design of the technological innovations at hand. Thus, creators of technological innovation occupy a particularly powerful, unique position in the relationship between the

infosphere and technological innovations. Though creators of technology are not the only moral agents who contribute to the application of technological innovations in the infosphere, their power over the nature of their creations clearly shows that not all causal factors in the consequences of technological innovations can be conceptualized as equal contributors. Better yet, creators of these systems have the unique power to define and re-define the very nature of their systems (and thus, the purposes, consequences, and target audience of these technologies).

To be more specific, creators of technological innovations have the power to change design features which pose a significant and known risk to user health, wellbeing, and their ability to operate as productive members of society. With intimate knowledge and understanding of their technology, developers can consider how or if certain elements of technological innovations can be removed, altered, or changed and the overall effect these changes will have on the function of the technological system at large. Furthermore, moral agents who create technological innovations have a unique understanding of the intention behind individual technological features and information technologies that is not easily perceptible to other moral agents or causal actors in the implementation of these larger systems. In this sense, creators of technological innovations have unique access to various elements of their creation process vital to assessing the extent to which they have given appropriate care to the ways in which their information technologies could (or have) impacted the

wellbeing of others. In simpler terms, creators of technological innovations hold unique power, influence, and agency in the more intimate and internal aspects of the technology while it is being constructed that can largely dictate whether the technological innovation was created in a morally responsible manner.

Conclusion

Overall, creators of technological innovations (including scientists and researchers) ought to be morally responsible for proactively and retrospectively considering the larger social and ethical consequences of their work. Considering how technological innovations enable new forms of life (that include new domains of moral responsibility), and their ability to (descriptively and normatively) influence our conceptions of moral and epistemic agency, technological innovations must be evaluated by their creators not just for their functionality. Instead, they must be also evaluated for their potential ethical consequences.

Some of these ethical consequences can be revealed by drawing on the works of Floridi (2013) and Winner (2014). By being embedded in our way of life (and thus, creating the infosphere), technological innovations hold significant power in determining our understanding of knowledge and thus, reality and value. Because of informational technologies' ability to rapidly re-ontologize the infosphere (and our inability to precisely and accurately identify which technologies will or will not have substantial moral or ethical consequences), rapid

technological innovation also comes with the risk of the passive re-ontologization of our physical world and those in it.

Because technological innovations can have such profound moral impacts, we must take on additional moral obligations to ensure the infosphere remains safe for its (human) inhabitants. As a result, I have argued that creators of informational technologies and technological systems must proceed with the utmost caution when developing technology. As a part of proceeding with the required caution, creators of technologies and informational technologies must be responsible for considering the larger consequences of their technological innovations, both before and after the technology has become integrated into the infosphere.

Although these consequences are unpredictable, it is important to consider that creators of technological innovation occupy a unique position and hold exceptional power and knowledge in altering their information technologies and understanding which aspects of an innovation ought to change or remain the same to prevent or minimize social consequences. Although creators of technologies are not the only causal factor for consequences of these innovations, their ability to largely determine and guide the intention of technological features (and the mechanical form in which they manifest) gives them considerable moral agency and responsibility not held by other, non-human causal factors or ontological actors.

Without holding innovators accountable for their technological creations, we will only continue to grow more unaware of the consequences of our technological actions until

we are forced to face them. Considering the increasing role technologies like ChatGPT are playing in our form of life, it is imperative we hold creators of technological innovations morally responsible for the impacts of their creations. Otherwise, we could risk losing our power as moral agents to normatively define and assign moral responsibility in the first place.

References

- Douglas, H. E. (2003). The moral responsibilities of scientists (tensions between autonomy and responsibility). *American Philosophical Quarterly (Oxford)*, 40(1), 59–68.
<https://www.jstor.org/stable/20010097>
- Duffy, Clare (2026). “Meta and YouTube found liable in social media addiction trial.” *CNN News*.
<https://edition.cnn.com/2026/03/25/tech/social-media-addiction-trial-jury-decision>
- Floridi, L. (2013). Ethics after the information revolution. In *The Ethics of Information* (pp. 1-18). Oxford Academic Books.
- Talbert, E. (2019). Beyond data collection: Ethical issues in minority research. *Ethics & Behavior*, 29(7), 531-546.
<https://doi.org/10.1080/10508422.2018.1531005>
- Talbert, M. (2025). *Moral Responsibility*. *The Stanford Encyclopedia of Philosophy*.
<https://plato.stanford.edu/archives/fall2025/entries/moral-responsibility/>
- Winner, L. (2014). Technologies as forms of life. *Ethics and Emerging Technologies*, 48–60. https://doi.org/10.1057/9781137349088_4

I Am, But Do I Think? On the Historical and Ongoing Devaluation of Philosophers as a Symptom of Thoughtlessness



"Why is he thinking? Did he lose his smartphone?"

Cartoon by Martha Campbell. © Martha Campbell / CartoonStock.

While the Western philosophical tradition has long characterized human beings as “rational animals” with thinking as our defining feature, “professional thinkers” have been undervalued, even devalued, since before Aristotle’s time and still today, particularly within capitalistic societies where thinking is generally not valued unless it directly produces profit. This paper aims to warn against the concrete dangers of this tendency towards thoughtlessness through a combined historical and future-oriented analysis. Drawing on Christopher Moore’s analysis of the early use of the term philosophos in Ancient Greece, I first trace the origins of this hostility toward philosophers and ‘philosophizing.’ I then trace this devaluation of intellectuals and intellectual activity to the development of a broader culture of thoughtlessness, drawing on Hannah Arendt’s account of the “banality of evil” as rooted in a lack of thinking that facilitates one’s fall into evil by undermining moral judgment. Applying Arendt’s insights on the dangers of thoughtlessness, I then examine how our growing reliance on generative AI risks further eroding our critical thinking and moral judgment skills by increasingly outsourcing thought itself in the name of ‘productivity gains.’ Based on this, I conclude with a defense of philosophy as a way of life that keeps us connected to what makes us distinctively human – our capacity to think – while keeping us away from thoughtlessness and the grave consequences it will have on society if more and more of us fall into it.

To be human is first and foremost to be capable of rational thought. Of course, to be human is also to feel, to love, to belong, to create, and all other meanings — if any — which we may attach to human existence to better situate and orient our existence within this world. Yet none of those characteristics are as distinctively human as our innate capacity for questioning and evaluating the world around us, for speaking and acting not just based on our desires, but on *deliberated* desire in the Aristotelian sense. In that regard, our rational capacities are what distinguishes us from other animals, fundamentally defining us as “rational animals.” Although this definition of human beings has been contested since then, the capacity to think rationally remains one, if not the most, defining aspect of human existence: we may not exercise it all the time and it can certainly be hampered by a multitude of factors including emotions, cognitive biases, and trauma, but the question to which I answer in the affirmative is not *are we perfectly or always rational?* like computers are designed to be, rather *are we rationally competent?* to a higher degree than other species. The “we” in question is also a key part of the answer: while most human characteristics and abilities are not entirely universal — the blind lack the ability to see, amputees lack certain body parts, diagnosed psychopaths lack the ability to feel empathy and guilt — every human being, by virtue of being human, has the capacity for rational thought.

The question then becomes whether, or to what extent, each one of “us” exercises it.¹

Yet do we, as a society, even value rational thinking all that much? If we did, then wouldn't those who dedicate themselves to thinking rationally and mastering its written and oral expression, that is philosophers, be some of the most valued members of society? Wouldn't we, following their example, seek to exercise and improve our rational thinking skills to our highest potential? Quite the contrary: today, those who write or teach philosophy 'for a living' have little to no influence beyond academia, and the more generative artificial intelligence (AI) develops, the more dependent we become on it to 'think' for us. Under a capitalist system, the underlying pressure seems to be that dedicating time and energy to thinking is only valuable, and thereby valued, when these efforts directly produce capital; and if an AI-powered chatbot can do the 'mental' work within

¹ Though the main argument of this paper will be largely based on the idea that our capacity for rational thought is fundamental and essential to our humanity, that is not to say that said capacity is inalienable or that it is necessary for us to have it in order to live: it can be gradually lost as a result of certain degenerative diseases like dementia, significantly reduced by a stroke or stripped away altogether by an accident leading to brain death (though in this case one loses all brain function beyond rational thought and would barely be considered a “living” human being anymore). Hence my emphasis on the “innate” character of this capacity, insofar as all human beings have it by virtue of being human, and one remains “human” even after having mostly lost it (e.g., due to severe dementia) by virtue of having had it before.

seconds, then why go through the effort myself? This logic, I will argue, is actively contributing to the societal problem of “thoughtlessness” as Hannah Arendt called it, of which the devaluation of philosophers, philosophy, and critical thinking as a whole is a symptom.

In examining the relationship between the phenomenon of “thoughtlessness” and this phenomenon of devaluation taken as its symptom, the latter will be termed *sophiaphobia* throughout this paper.² From the Greek words σοφία (sophia) for “wisdom” and Φόβος (phobos) denoting “aversion” or “fear,” *sophiaphobia* refers to a form of anti-intellectualism directed specifically towards the intellectual pursuit characterized by philosophers, that is the search for “wisdom” through independent, critical thought. Such aversion to the activity of thinking itself may stem from one’s prioritization of profit-seeking instead, from one’s fear of engaging in the strenuous effort that profound thinking represents or from one’s fear of the disruption which “professional thinkers” can bring to society (with Socrates as a quintessential example). Whatever the cause of such an attitude in each case, this paper will focus on its consequences: examining what happens when *sophiaphobes* dismiss, mock, or devalue those who dedicate themselves to the activity of thinking, as well as when this devaluation becomes so entrenched in a certain

² Special thanks to my editor, Thomas, for suggesting that I coin such a term as a more context-specific alternative to “anti-intellectualism.”

system or society that it manifests itself in the thoughtless behavior of its members.

Given that it can manifest itself in various forms, sophiaphobia can be explored through different angles. In this paper, I will approach it from a combined historical and future-oriented perspective, beginning with the roots of hostility and mockery towards philosophers in fourth century BCE Athens, around the time when the term *philosophos* was coined as *philosophia* eventually formed into a discipline. The purpose of this historical analysis, based on Christopher Moore's *Calling Philosophers Names: On the Origin of a Discipline* (2019), will be to show that dismissal and mockery of philosophically-inclined individuals is not just a modern-day phenomenon, but one as old as Western philosophy itself (I). Following a brief discussion of how the public perception and societal influence of philosophers have evolved over time, I will turn to a deeper problem, beyond Ancient Greece, tied to the dismissal and devaluation of that which philosophers practice and exemplify that is the activity of thinking itself. In examining what such tendency to de-prioritize critical, independent thought can concretely do to society as a whole, I will rely on Hannah Arendt's analysis of "thoughtlessness" within the context of twentieth-century totalitarianism (II). Based on her analysis of this "absence of thinking" as the root of a particular type of evil which she saw

embodied by Nazi official Adolf Eichmann,³ I will explore the destructive consequences to which thoughtlessness led in the twentieth century, before turning to the consequences that it may have today amid technological advancements and our overreliance on certain forms of technology **(III)**. Keeping in mind Arendt's warnings, I will aim to show why the delegation of thinking to generative AI is an ongoing manifestation of the problem of thoughtlessness and why every member of society – not just philosophers – should actively resist it today. The paper will then conclude with a return to the origin: a Platonic call for more widespread engagement with philosophy as one possible approach to targeting the problem of thoughtlessness at its root and thus staying clear of its potentially destructive consequences – both for oneself and for wider society.

Importantly, the following examination of sophiophobic attitudes aims to establish neither a causal link nor a direct historical connection between such attitudes and the phenomena of thoughtlessness examined in these three distinct historical periods. Rather, the aim of this paper is to show, based on three

³ “[...] no theory or doctrine but something quite factual, the phenomenon of evil deeds, committed on a gigantic scale, which could not be traced to any particularity of wickedness, pathology, or ideological conviction in the doer, whose only personal distinction was a perhaps extraordinary shallowness” – a type of evil which she called the “banality of evil.” More on this in section II – On the Problem of Thoughtlessness: A Twentieth Century Perspective.”

case studies, how sophiaphobia can go hand-in-hand with a certain embrace of thoughtlessness, insofar as the latter is the basis of the “unexamined life” which sophiaphobes seek to live and entrench in wider society.

“Foolish,” “Sage-Wannabe,” “Teacher of Misery:” The Emergence of *Philosophos* as a Scornful Label in Ancient Greece

The fact that Athens was a hub of philosophy where quintessential philosophers such as Socrates and Plato gained recognition does not mean that every Athenian was a philosopher, nor that all Athenians were favorable toward these members of its society. In fact, as Christopher Moore argues based on textual and archaeological evidence, the label *philosophos* initially carried a pejorative connotation. Only years later was it gradually redeemed and ultimately reappropriated by those who, united by a shared “love of wisdom,” came to form their own discipline under that name: thus emerged *philosophia*, from the Ancient Greek words φίλος (*philos*) for “love” and σοφία (*sophia*) for “wisdom.”⁴ In that way, according to Moore, Athenian thinkers whom we now remember as some of the first “philosophers” came to regard themselves as such

⁴ Christopher Moore. *Calling Philosophers Names: On the Origin of a Discipline*. Princeton: University Press, 2020, 1.

through a process of reappropriation and ennoblement of a slur – similar to how initially pejorative labels such as “queer” and “Shaker” have been gradually reappropriated and redefined by the targeted groups. Thus, as we will see in this first section, *philosophia* was born within a context where such dedication to critical thinking was not valued or even welcome, especially not when it came anywhere near “corrupting” the elite – a testament to the low regard in which philosophers were held within a society dominated by *sophiaphobia*, which I will argue resembles our societal predicament today.

Given the etymology of the word “philosophy,” we tend to define philosophers simply as “lovers of wisdom,” glossing this as “lacking-but-striving-for-wisdom.”⁵ Yet Moore’s historical reconstruction reveals that this was not what people originally meant by “philosopher,” or **φιλόσοφος** (*philosophos*) in its original Greek form. Back when the term was coined around the sixth and early fifth centuries BCE, the *phil-* (**φιλ-**) prefix which it contains did not designate such “lack:” on the

⁵ Moore, 66. Though Moore does not further elaborate on the terms of this “gloss,” I would clarify that “lack” of wisdom does not entail lack of knowledge or lack of intelligence: one may be considered a “philosopher” by virtue of having already reached a certain intellectual status through demonstrated knowledge and intelligence, while still striving towards the attainment of a more complete form of “wisdom.” In that sense, this “glossed” definition of what constitutes a philosopher implies both lifelong intellectual striving and a certain humility maintained throughout.

contrary, it implied clear or regular presence of the second element's referent.⁶ Based on this finding by scholar Walter Burkert along with further evidence, Moore argues that *phil*-prefixed names (of which Greek has hundreds) were originally pejorative "name-calling names," calling out "excessive activity" related to a social practice designated by the word's second element.⁷ This suggests that *phil*- was generally attached to nouns to form labels used not for congratulations or approval, nor as self-descriptions, but as *other*-applied epithets carrying a "bemused, skeptical or even censorious judgment."⁸ Based on this pattern, Moore's thesis is that calling attention to someone's relation to *sophos* by attaching a *phil*- prefix to it implied that the

⁶ Moore, 66.

⁷ Moore, 6. To take one illustrative example among the many cited by Moore, consider **φιλαίτιος** (*phil* + "legal motions"), which refers to "excessive activity" in lawsuits. In itself, filing legal motions is a normal activity – which Moore suggests is precisely what makes the name pejorative rather than laudatory: "since one's relation to legal motions usually is not a matter of discussion, calling attention to it by creating a name for it implies that the relation is not merely notable: it is worrisome" (Moore, 85). For a more detailed account, see Chapter 3.

⁸ Moore, 84. To add on to Moore's historical analysis, consider how this ancient Greek pattern of insult is still used in modern English, with "[slur]-lover" used, often with racist intent, as an insult towards someone considered excessively fond or protective of the designated group.

relation was ridiculous or problematic in the eyes of the name-caller.⁹

Moore then supports his thesis by exploring the specific connotation of *sophos* as the second element of the *phil-* prefixed name in question. Towards the end of the sixth century BCE, the term referred specifically to the epistemic authority held by the *sophoi* or “sages” – socially elite and influential advisers typified by the “Seven Sophoi” or Seven Wise Men of Ancient Greece. Given this and the aforementioned pattern, Moore suggests that calling someone *philosophos* back then denoted a tendency to act like those respected *sophoi* through “excessive” study and advising¹⁰ while claiming undue epistemic authority.¹¹ Indeed, one traditionally did not train to become *sophos* as a young man;¹² only at a mature age could such a title be earned.¹³ Thus, from an external perspective, young men studying to become *sophos* – which those engaged in deep conversation about life,

⁹ Moore, 6.

¹⁰ Moore, 6.

¹¹ Moore, 93.

¹² It should be noted that, although they constituted a minority and gained pushback from the political elite, these “young men” were typically citizens, meaning that they were themselves part of a certain elite relative to the whole of Athenian society. This social status gave them access to education and other rights that allowed them to circulate in Athenian public life as “wisdom-lovers,” a privilege not granted to female citizens nor to non-citizens altogether.

¹³ Moore, 105.

death, and justice were seen as doing – was an “odd, even risible idea” at best.¹⁴ At worst, it was deemed “hubristic” to try to “shortcut” one’s way to *sophos* status:¹⁵ while most people respected the sages, these wisdom-seeking young men went further – too far – by trying to emulate them. Such deviation from the norm is what earned those who would later be considered “philosophers” a *phil-* prefixed label, in line with the classic social tendency to label those who dare to be different. Thus, following this account, *philosophos* was likely coined as a result of the popular impulse to mock or condemn a group of people for their unreasonably excessive “love of wisdom” and illegitimate aspirations to sagehood.¹⁶

Given the reputation of those to whom it was attributed, *philosophos* was not a term of aspiration, acclaim, or self-identification in the sixth and fifth centuries BCE.¹⁷ It wasn’t until the fourth century BCE that the term started to be redefined, mainly through the figure of Socrates – a man considered the living embodiment of wisdom-loving and

¹⁴ As Moore argues in Chapter 4, these young men were likely the Pythagoreans, followers of Pythagoras (one of the first self-proclaimed *philosophos*), who traveled around the Magna Graeca enough to be seen by many people – not all of whom would have appreciated or understood their studious activity (Moore, 108).

¹⁵ Moore, 105.

¹⁶ Moore, 105-106.

¹⁷ Moore, 45.

wisdom-seeking.¹⁸ Indeed, Socrates was regarded as a *philosophos* within his lifetime for his commitment to practicing and spreading *philosophia* through questioning everything and everyone around him, including the political establishment. Such intellectual audacity earned him a reputation as a “gadfly,” which eventually led to his execution at a time when public opinion around philosophers (*philosophoi*) was already condemnatory: incomprehension and mockery towards Socrates’s unconventional ways turned into animosity when the political elite, seeing how influential he was among their children, set out to stop him from “corrupting” the future rulers of Athens.¹⁹ This accusation of ill intent on the basis of philosophical activities perceived as dangerous for the public may be the most extreme manifestation of sophiaphobia directed against the first Western philosophers. Following Socrates’s execution, his former students Plato and Xenophon sought to redeem his image by making him a central figure of their writings: though they rarely had Socrates’s character explicitly call himself or get called *philosophos* (likely given their awareness of the term’s enduringly pejorative connotation²⁰), they gradually redefined what it meant to be one by portraying Socrates as the symbol of critical thinking, ethical inquiry, and

¹⁸ Moore, 28.

¹⁹ Moore, 166.

²⁰ Moore, 158.

intellectual courage that we still remember and study today – especially as modern-day philosophers who, for many of us, strive to live by his example amid various forms of pushback from *sophiaphobes*.

Among ancient writings about Socrates and his philosophical activities, Plato's dialogues most significantly contributed to the redemption and revaluation of the term *philosophos* by showing how *philosophia*-style conversations can lead to virtue and flourishing.²¹ A central component of this redemptive project was his depiction of Socrates not just as a symbol, but as an "ideal."²² In Plato's *Apology*, Socrates denies the charges brought against him and exposes his accusers' ignorance around what it really is to 'philosophize,' that is "'to live [...] examining [one]self and others.'" In the *Charmides*, Socrates presents *philosophia* as a conversation towards self-improvement.²³ In the *Phaedrus*, while dialoguing with his interlocutor Lysias, Socrates explicitly gives a laudatory connotation to *philosophos*, describing the latter as someone with the ability to make speeches based on knowledge rather than mere opinion and to defend those speeches while recognizing their limited value.²⁴ Here, the condemnation is not just

²¹ Moore, 29.

²² Moore, 190.

²³ Moore, 223.

²⁴ Moore, 231.

countered but reversed against non-*philosophoi* like Lysias, depicting them as the ones at fault while suggesting that one should aspire to be like the *philosophoi* instead.

As Moore notes, this move reflects a second dimension of Plato's redemptive project: beyond just reappropriating a previously scorned activity for the purpose of redeeming its practitioners, he sought to ennoble *philosophia* as a "beneficial conversational-educational activity"²⁵ for *all* – subliminally telling each of his readers, "Whoever you are, and whatever your current goals, you *should* philosophize!"²⁶ With this two-fold "mission" advanced in his dialogues and spread through the latter's timeless popularity, Plato can be credited not just with contributing to the revaluation of *philosophos* as a proudly embraced identity marker among those who already identified with it more quietly, but also with expanding *philosophia* to a wider audience beyond said group of self-proclaimed *philosophoi*. Accepting the charge that Socrates "philosophized," Plato set out to advance the *universal* value of "philosophizing" by showing why Socrates's example *should* be followed rather than scorned. As Moore puts it, a practice beneficial to some *as humans* – in this case, "talking about fundamental issues necessary for improving oneself and others"²⁷ – is a practice

²⁵ Moore, 257.

²⁶ Moore, 259 (italics added).

²⁷ Moore, 257.

actually beneficial to all *as humans*;²⁸ by making this idiosyncratic practice “normative,” the practice “loses its idiosyncrasy.” Thus, the call to philosophize becomes normative not just for some, but for all²⁹ – including the sophiaphobes.

Did Plato’s reappropriation and revaluation of *philosophia* succeed in the face of Athenian *sophiaphobia*? Though his redemptive project may not have convinced all of his contemporaries, the fact that Socrates’s legacy has since been that of an example rather than a malicious influence or unreasonable “sage-wannabe” may be taken as a measure of success. Beyond redeeming Socrates’s reputation and the *philosophos* label altogether, Plato is generally thought to have “spearheaded” the formation of Western philosophy as we know it, with “self-consciously philosophical” practitioners.³⁰ The discipline has evolved quite a bit since the fourth century BCE, with considerable influence exerted over politics, social institutions, and intellectual debates over the centuries: Aquinas helped

²⁸ Moore, 258 (italics added).

²⁹ Moore, 259. I follow Moore’s Universalist interpretation here, while acknowledging that it may not reflect exactly *who* Plato himself had in mind when suggesting that one should aspire to be like the *philosophoi*. In fact, it should be noted that he believed not everyone was born with the ability to become a “philosopher,” at least not in the way he presented the title in the *Republic*. At the same time, he did allow not just philosopher kings, but also philosopher queens to rule his ideal *polis*, which legitimizes somewhat broadening the normative scope of a Platonic call to philosophize.

³⁰ Moore, 190.

shape the foundations of Christian doctrine, Machiavelli's advice continues to influence rulers to this day (such as President Macron), Rousseau's ideas inspired those who went on to lead the French Revolution just as Locke directly influenced those who founded the United States, Beauvoir's *The Second Sex* became the blueprint of second wave feminism – to name just a few in the Western tradition.

Yet despite Plato's efforts and the subsequent contributions of self-proclaimed philosophers, it appears as though we have fallen back into an era of dismissal (at best) and devaluation (at worst) of philosophy as both a discipline and a lifestyle – comparable to the pre-Socratic period described by Moore. If Plato's vision of philosopher kings and queens in the *Republic* was already idealistic at the time, it seems almost inconceivable today in the absence of any philosopher as culturally prominent and socially influential as Plato and Socrates were. Indeed, the fact that Socrates was sentenced to death is but a testament to the amount of influence he held; today, how many philosophers are sitting on death row or getting publicly executed for "corrupting" a wide audience? Fortunately for current-day philosophers, very few, if any.³¹ Today, those who identify as philosophers generally hold

³¹ In fact, if one looks through famous philosophers' causes of death over the centuries, some historical tendencies come to light: several ancient philosophers were murdered or executed, several Renaissance philosophers as well, and several Modern

little influence beyond academia, while those who study philosophy are often met with incomprehension, mockery or genuine concern: that this degree is useless and will leave us unemployed, that it is a waste of money... And should we decide to pursue philosophy beyond our studies, then we might be met with variants of Callicles's criticism in *Gorgias*: that philosophizing has its charms for the youth but makes adults incapable of facing the hard facts of political life, that we ought to 'act' instead...³²

I would like to argue that this tendency to dismiss and devalue *philosophia*, its students and its practitioners is a symptom of a larger societal problem, one that we might qualify as disinterest towards the activity of thinking itself. One could attribute this to our economic system, wherein an activity only has value if it can be quantified in terms of income or profit. Insofar as studying philosophy teaches students how to make a

philosophers too but only under authoritarian regimes that espoused anti-intellectualism (e.g., Gustav Shpet in the USSR, Kurt Grelling in Nazi Germany). Aside from that, the second leading cause of death among famous philosophers since the end of the Cold War, after disease, has been suicide. I mention this not as commentary on the number of suicides among modern philosophers but as one possible measure of the decline in philosophers' influence over time, insofar as the execution of philosophers in the past usually meant that they were considered influential enough to pose a legitimate threat to political power through their ideas and status in society. (Cf. "Deaths of Philosophers." *Wikipedia*. Last modified January 2026.)

³² As paraphrased by Moore, 134.

‘good life’ over how to make a ‘good living,’ a philosophy degree does not lead to a stable income as directly as other degrees might – which, to those who see philosophy primarily as a college degree or class, might make philosophical practice seem useless altogether. This “commercial” view of *philosophia* actually predates capitalism: back in Xenophon’s *Memorabilia*, one scene features the sophist Antiphon mocking Socrates’s poverty and painting the philosophical lifestyle as ridiculous. “I thought that those philosophizing had to end up flourishing more,” he remarks sarcastically. “But you [Socrates] seem to me to have won the opposite from *philosophia*... After all, you are poor and do not charge tuition...”³³ According to Antiphon (who fruitfully charged his own students³⁴), philosophy is reducible to

³³ Xen. *Mem.* 1.6.2–3. Quoted in Moore, 180.

³⁴ Beyond the characters of Socrates and Antiphon, this scene displays a confrontation between two approaches to ancient Greek intellectualism which they respectively embodied: *philosophoi* on one hand, sophists on the other. Though some argue that the distinction was less clear-cut in reality, Plato sharply distinguished philosophers from sophists, of whom he was particularly critical (namely for their profit-seeking approach to teaching). Sophists had their own criticisms in return, at least as presented in Plato’s and Xenophon’s writings. Yet it should be specified that this does not fall under what I have termed “sophiaphobia:” while involving scorn towards philosophical activity, criticism of philosophers by sophists only targets the manner of execution of such activity, not the activity itself. In contrast, “sophiaphobia” involves a certain aversion or at least rejection of the very activity of thinking by those who (unlike sophists) do not identify as ‘intellectuals’ themselves.

a pedagogical profession and can only make one “flourish” through payments from students. Yet Socrates stood as a living rebuttal to both certitudes: because he spread and sought wisdom for free and did not construe “flourishing” in material terms, he could not be valued as any more than a “teacher of misery” by people like Antiphon.³⁵ This is but one manifestation of the fundamental tension between wisdom-seeking and money-orientedness: as Socrates was misunderstood and scorned back then, anyone who dedicates themselves to prolonged thinking without directly seeking financial compensation out of it, and while not necessarily having other funds to rely on either, may well be misunderstood and scorned today.

Thus, disinterest towards the activity of thinking – when not directly for the sake of compensation – might stem from our society’s money-orientedness. While this discussion of causes could be pursued further, I would like to focus the rest of this paper on a discussion of consequences instead. Indeed, the stakes of this problem extend far beyond whether philosophers get the respect and attention which I think they (or we) deserve: that with which we must reckon more universally is what the societal devaluation or de-prioritization of the kind of activity exemplified by philosophers – thinking critically for oneself – can do to society as a whole. We saw the consequences of this

³⁵ Xen. *Mem.* 1.6.2–3. Quoted in Moore, 180.

phenomenon in ancient Greece, where politically upheld sophiaphobia first manifested itself as scorn towards those “excessively” dedicated to philosophical practice and later culminated in the killing of Socrates for having dared to philosophize, or question, too far. Such contempt and fear of wisdom-seeking can be regarded as a certain embrace of the opposite activity: underusing one’s critical thinking skills and embracing “thoughtlessness.” As I will show in the following section based on Hannah Arendt’s analysis of this phenomenon, “thoughtlessness” – like sophiaphobia – can be deadly when embraced at a political or wider societal level. It therefore must not be underestimated, especially as society becomes more and more reliant on technology and less on human capacities.

On The Problem of Thoughtlessness: A Twentieth Century Perspective

If, as I posited earlier, our capacity for rational thinking is what makes us distinctly human, it follows that the less we exercise that capacity, the more we alienate ourselves from our humanity. In other words, the less we think, the more we reduce ourselves to acting like non-human animals or machines – neither of which have a developed moral compass, if any. Indeed, “thoughtlessness” is not just an intellectual problem but an inherently moral one: beyond blocking one’s engagement with what we have characterized as a universally beneficial intellectual practice – that is “talking about fundamental issues

necessary for improving oneself and others,”³⁶ as Moore puts it – embracing thoughtlessness also means letting go of one’s crucial capacity for moral reasoning. In Kantian terms, this necessary connection between the capacity for critical thinking and the capacity for moral reasoning may be articulated in terms of “autonomy” and “moral agency.” Autonomy, as Kant understood it, refers to the capacity of rational agents to make their own choices based on reason, free from external influences or coercion. Far from an abstract philosophical concept, autonomy is a cornerstone of Kant’s timelessly applicable moral framework: insofar as to act autonomously is to exercise one’s capacity for rational thought and moral judgement in accordance with moral laws that one rationally ‘legislates’ *for oneself*, autonomy is essential for moral agency.³⁷ In other words, if I am to consistently make sound moral decisions, then I must actively exercise my rational capacities. It follows by *modus tollens* that if I stop exercising my rational capacities, then my capacity to consistently make sound moral decisions will thereby be eroded – from which the most unreflectively immoral actions may follow.

³⁶ Moore, 257.

³⁷ This reconstruction of the connection between autonomy and moral agency in Kant’s moral framework was borrowed from “The Influence of Kantian Philosophy on Today’s World.” *Philosophy Nest*, November 18, 2025.

While such conclusions may be dismissed as merely hypothetical or alarmist philosophical musings, Hannah Arendt gives us all the more reason to take them seriously. As both a German-born Jewish woman forced to flee Nazi Germany and a political philosopher significantly influenced by Kant, Arendt dedicated much of her life to grappling with the crucial political events of her time and demonstrating their impact on our moral judgement — namely with regards to the horrors of totalitarianism.³⁸ Within this context, she paid particular attention to the nature of *evil*, initially describing the evil committed by totalitarian regimes as “radical,” using Kant’s expression.³⁹ That is, until she distanced herself from this Kantian analysis after witnessing the trial of Nazi official Adolf Eichmann in Jerusalem, to which she was drawn out of curiosity as to what someone having committed “radical evil” might look like: there as a reporter, what she found was not a “monster” as expected, but an “ordinary, commonplace” man displaying “a curious, quite authentic inability to *think*.”⁴⁰ Such “total absence

³⁸ Tatjana Tömmel and Maurizio Passerin d’Entrèves. “Hannah Arendt.” *The Stanford Encyclopedia of Philosophy*, Spring 2025 ed., ed. Edward N. Zalta and Uri Nodelman.

³⁹ Tömmel and Passerin d’Entrèves. “Hannah Arendt.”

⁴⁰ Hannah Arendt. “Thinking and Moral Considerations: A Lecture.” *Social Research* 38, no. 3 (1971): 417–446, 417.

of thinking,"⁴¹ she later argued, lies at the root of evil – hence her subsequent conceptualization of “thoughtlessness” as a *moral* problem.

Thus, it was the Eichmann trial that not only redirected Arendt’s understanding of the nature of evil but also sparked her interest in the phenomenon of thinking,⁴² laying the ground for her later writings on the moral significance of thought and judgement. What struck her while watching this man who had managed the deportation of Jews to death camps for years was the lack of “wickedness, pathology or ideological conviction” displayed during his trial: the deeds were monstrous, but the doer apparently was not.⁴³ In fact, his “only personal distinction was a perhaps extraordinary shallowness”⁴⁴ displayed through his reliance on “clichés” and “stock phrases” as well as his “adherence to conventional, standardized codes of expression and conduct” throughout the trial.⁴⁵ Absent demonstrated characteristics of an “evil” character or clear affirmation of hatred toward Jews, the only explanation Arendt found for Eichmann’s evil deeds, aside from his “law-abiding” devotion to

⁴¹ Arendt, “Thinking and Moral Considerations: A Lecture,” 418.

⁴² Hannah Arendt. *The Life of the Mind*. Harcourt Inc., 1981, 6.

⁴³ Arendt, “Thinking and Moral Considerations: A Lecture,” 417.

⁴⁴ Ibid.

⁴⁵ Arendt, “Thinking and Moral Considerations: A Lecture,” 418.

“the Führer,” was his “thoughtlessness”⁴⁶ – a characteristic displayed not just during the trial, but just as much during the preceding police examination and during his years as a Nazi official prior.⁴⁷ Thus, Arendt claimed, “it was sheer thoughtlessness – something by no means identical with stupidity – that predisposed [Eichmann] to become one of the greatest criminals of that period.”⁴⁸ Following the trial, she coined the concept of “banality of evil” to account for the discordance between Eichmann’s “moral and intellectual shallowness”⁴⁹ on one hand, and his evil deeds on the other.

Before we build off of Arendt’s conclusions, it would be intellectually dishonest not to nuance the portrait of Eichmann which she proposed. Granted, her judgements of his character and motives were based on a limited scope of first-hand observations, whose only object was the questionably honest behavior and statements of a man on trial. Recognizing the contingency of her claims, Arendt explicitly noted that “the banality of evil” which she derived from Eichmann was meant

⁴⁶ Hannah Arendt, *Eichmann in Jerusalem: A Report on the Banality of Evil*, 149.

⁴⁷ Arendt, “Thinking and Moral Considerations: A Lecture,” 417.

⁴⁸ Arendt, *Eichmann in Jerusalem: A Report on the Banality of Evil*, 288. As clarified later in her introduction to *The Life of the Mind*, what Arendt means by “stupidity” here is the “inability to comprehend.”

⁴⁹ Arendt, *Eichmann in Jerusalem: A Report on the Banality of Evil*, Introduction by Amos Elon, xiii.

to denote not a general “doctrine” or “theory” of the nature of evil, but only what she had observed as “undeniably factual” about Eichmann’s personality.⁵⁰ Yet evidence released years after her death reveals that her assertions on Eichmann’s personality were actually undeniably mistaken: while she saw only “shallowness” and Eichmann himself claimed that he had only followed orders based on a Kantian definition of duty,⁵¹ recently discovered tapes from 1957 reveal that he was actually a committed ideologue. Dubbed the Eichmann tapes, these recordings of an interview with a Nazi journalist in Argentina, four years before the trial, feature Eichmann bragging about his role in the Final Solution and stating that he would have been “satisfied” if “millions” more Jews had been “exterminated.”⁵² Given this, the idea that only Eichmann’s “thoughtlessness” could explain his evil deeds cannot be reasonably upheld anymore.

What do these revelations entail for Arendt’s concept of “banality of evil,” given its tie to Eichmann’s example? Some

⁵⁰ Arendt, *The Life of the Mind*, 4.

⁵¹ Arendt, *Eichmann in Jerusalem: A Report on the Banality of Evil*, 136. For Arendt’s analysis and deconstruction of Eichmann’s stated adherence to “Kantian precepts,” see VIII: Duties of a Law-Abiding Citizen.

⁵² For more on these revelations, see Chakrabarti, Meghna and Kotsonis, Stefano. “The Eichmann Tapes and the Comforting Myth of the ‘Banality of Evil.’” *On Point* (WBUR), July 15, 2022.

may want to reject her insights altogether, like many did, especially among Jewish circles, amid significant controversy following the publication of *Eichmann in Jerusalem*,⁵³ in fact, some of her critics at the time may have been right to accuse her of portraying an anti-Semitic monster as a mere bureaucrat. Along those lines, the concept of the banality of evil may become nothing more than a “comforting myth.”⁵⁴ Others may want to retain the concept insofar as it can still be applied to lower-level bureaucrats unreflectively obeying orders. Along these lines, my position is that we should continue to uphold Arendt’s “banality of evil” as a timelessly relevant conceptual framework to help us identify, understand, and ultimately prevent gradual descent into thoughtlessness and immorality amid temptations to stop thinking independently from the systems and institutions⁵⁵ in which we are embedded. Indeed, however factually misled they

⁵³ Tömmel and Passerin d’Entrèves. “Hannah Arendt.”

⁵⁴ Chakrabarti and Kotsonis, “The Eichmann Tapes and the Comforting Myth of the ‘Banality of Evil.’”

⁵⁵ By “institutions,” I refer not just to physical organizations founded for a given collective purpose (including educational “institutions” like York University), but also to established social practices and customs that structure our lives and behavior (e.g., family, marriage, legal systems). Taken in this broader sense, “thinking independently from institutions” means exercising our capacity to critically evaluate, challenge, and form moral judgements (i.e., “philosophize”) on questionable matters that may be presented as “just the way it is” in these institutional structures (e.g., certain marriage norms, certain unjust laws), and doing so despite sophiophobic attempts to quiet our questions.

might have been, Arendt's observations during the trial led her to a conclusion extending beyond Eichmann's case, even beyond the context of totalitarianism which she studied: that any "ordinary" human being who simply stops thinking critically about what he or she is doing can come to carry out the worst forms of evil without even realizing it, or without the moral reasoning skills left to care. Hence the question she posed a few years later: "Might the problem of good and evil, our faculty for telling right from wrong, be connected with our faculty of thought?"⁵⁶ The answer, with or without Eichmann's case as corroboration, remains yes.

By drawing this connection between the faculty of thought and the problem of evil, Arendt's analysis of one criminal became a universal warning about the danger, destruction, and death that may eventually result from something as seemingly innocent as *doing* without *thinking*: that is, gradually weakening our rational capacities by living unreflectively, and living unreflectively as a result of weakened rational capacities. Considering the level of immorality that we may easily attain through this vicious circle, the previously mentioned call to philosophize appears as urgent as ever – hence the enduring public utility of philosophical practice. Such practice may take various forms – from Socrates dialoguing in the *agora* to Arendt critically reporting on a courtroom trial to Moore and myself

⁵⁶ Arendt, *The Life of the Mind*, 5.

building on their contributions in writing – as long as it involves actively exercising one’s critical thinking and moral judgment skills against whatever pushback may come from sophiaphobes. In pursuing this lifestyle, philosophers may be regarded as both examples and advisers: similar to how Moore interpreted Plato’s redemptive project as including the injunction, “Whoever you are, and whatever your current goals, you should philosophize!”⁵⁷ we may interpret Arendt’s account of thoughtlessness as telling us, “Whoever you are, and whatever your current goals, do not stop thinking!” Indeed, the final (posthumously published) book she wrote suggests this injunction in the form of an open question: “Could the activity of thinking as such, the habit of examining whatever happens to come to pass [...] be among the conditions that make men abstain from evil-doing or even actually ‘condition’ them against it?”⁵⁸ As she develops in her analysis, the answer, again, is yes. In other words, if the absence of thinking is a possible cause of evil-doing, then active thinking is one of the simplest, yet most fundamental safeguards against the latter.

And yet as simple as active thinking may seem compared to other “activities,” here we are in need of a reminder as to why we not only should, but *must* give it more of our time and attention: to form healthy communities of reflective, perpetually

⁵⁷ Moore, 259.

⁵⁸ Arendt, *The Life of the Mind*, 5.

self-improving individuals, but more consequentially to protect human society from the possibly deadly consequences of “thoughtlessness” writ large – of which totalitarianism, upheld largely by “thoughtless” low-level bureaucrats, is one example. The stakes are high, yet we still tend to let “the absence of thinking” take over: it is “so ordinary an experience in our everyday life,” as Arendt observes, “where we have hardly the time, let alone the inclination, to *stop* and think.”⁵⁹ In the following and final section, I will draw on this observation to explore how such “ordinary” thoughtlessness manifests itself today, a few decades after Arendt’s writings, within an increasingly technologically advanced context. Focusing on the development of generative AI and our growing reliance on it, I will argue that we are increasingly not just weakening, but losing the inclination to “*stop* and think” at the core of a reflective, moral life. Keeping in mind Arendt’s warnings about “thoughtlessness,” I will then aim to show why this *delegation* of thinking, beyond the “disinterest” considered earlier, is a problem that every member of society – not just philosophers – should care about and actively resist today. The paper will then conclude with a return to the origin, presenting a Platonic call for more widespread engagement with philosophy as one solution to the problem at hand.

⁵⁹ Arendt, *The Life of the Mind*, 4.

On The Problem of Thoughtlessness: A Twenty-First Century Perspective

*"While I like to think of myself as a decent writer, at this point, I often believe that LLMs can explain things better than I can."*⁶⁰

At first glance, this comment found on an online forum may seem like nothing more than the commenter's humble perception of their own writing. Yet, it struck me as something more harmful than that. Judging by their active participation on this forum, it can be reasonably inferred that said commenter is a relatively knowledgeable person who knows how to write and does not *need* "LLMs" (as in Large Language Models, the basis of chatbots like ChatGPT⁶¹) to formulate their thoughts for them. Yet, judging by this comment, they seem to have convinced themselves that they do – an expression of our generation's increasing surrender to generative AI. As I will argue, this phenomenon runs counter to the previously established call to philosophize, or simply to think for *oneself*, and is thereby contributing both to thoughtlessness and to an increasingly normalized form of sphiaphobia.

⁶⁰ From the comment section of William MacAskill's "EA: A View from the Inside" posted November 17, 2020, on the *Effective Altruism Forum*.

⁶¹ Lisa Lacy, "LLMs and AI Aren't the Same: Everything You Should Know About What's Behind Chatbots," *CNET*, May 31, 2025.

Generative AI can be defined as a type of artificial intelligence that *generates* new content such as text, images and videos in response to inputted prompts from a user.⁶² Rather than being trained only to make predictions based on a specific dataset (which is what AI originally came down to), the particularity of generative AI is that such models are trained to *create* new data resembling that on which they were trained.⁶³ This generative aspect is central to my argument insofar as it mimics the human faculty of thought: one which distinguishes itself from the thought of other beings by virtue of the distinctly human capacities to judge, evaluate, and then draw novel conclusions based on that ‘data.’ In Kantian terms, such cognitive activity is permitted by the “power of judgment” (*Urteilstkraft*) or “faculty of judging” (*Vermögen zu urteilen*): this spontaneous and innate cognitive capacity is that through which one makes judgments that are essentially “propositional cognitions;” from this it follows that rational human beings are “propositional animals.”⁶⁴ Only humans have such capacities in Kant’s view; yet what about those machine-learning models which now *present* themselves as equal to humans in this

⁶² “What Is Generative AI?,” *Coursera*. Updated on April 18, 2026.

⁶³ Adam Zewe, “Explained: Generative AI,” *MIT News*, November 9, 2023.

⁶⁴ Robert Hanna, “Kant’s Theory of Judgment,” *The Stanford Encyclopedia of Philosophy*, Spring 2022 ed., ed. Edward N. Zalta. This theory of cognition is mainly developed in Kant’s *Critique of Pure Reason*.

propositional regard? Since the “generative AI boom”⁶⁵ which brought chatbots such as ChatGPT into the hands of anyone with access to a browser (and a VPN where needed), we now have a third entity to turn to other than ourselves and other humans when it comes to making “propositions,” that is the content of judgments.⁶⁶ Although, for Kant, propositions can only be “generated” through psychological processes, thus by “judging animals” with the same cognitive makeup, AI chatbots can now “generate” them just as well – perhaps even better than oneself, following the aforementioned commenter’s reasoning. Given this, and considering how easy it is to use these chatbots compared to how long and mentally strenuous it can be to draw one’s own conclusions, and how others may not always be available to help, the temptation to resort to generative AI to judge and formulate ideas *for us* is certainly understandable. Yet the more we let it think for us, the less we think for ourselves – which is where I think the harm comes in.

⁶⁵ Zewe, “Explained: Generative AI.”

⁶⁶ As Hanna explains based on selections from the *Critique of Pure Reason*, “a proposition is the logically well-formed and semantically well-composed, truth-valued, unified objective representational content of a judgment; and more generally it is “what is judged” in the act of putting forward any sort of rational claim about the world [...] In this way judgments for Kant are *essentially* propositional cognitions, in that the primary function of the faculty of judgment is just to generate these logically well-formed, semantically well-composed, truth-valued, intersubjectively shareable, rationally communicable, unified objective representational contents.”

To be clear, this is not to say that generative AI is *inherently* harmful; like any tool, it is only as harmful as the user makes it. Moreover, what I see as harmful is not our reliance on it to *help* us in our thinking: just as any tool may help us perform a given task or activity, chatbots can help us better understand certain ideas and test our own ideas through a simulation of dialogue, while helping us perform certain inherently unreflective tasks more efficiently (such as gathering references under the same citation style). What I do see as harmful is our reliance on its ‘judgment’ *rather than* on our own when it comes to generating ‘new’ ideas or content.⁶⁷ In that capacity, generative AI becomes more than just a tool or piece of “equipment” in the Heideggerian sense – hence why it must be distinguished from other technological advancements:⁶⁸ when

⁶⁷ While proper consideration of the following points lies beyond the scope of this paper, it should be noted that there are reasons to regard generative AI as “harmful” *by design*: one being its irreversible environmental impact, another being its designers’ wide-scale theft of intellectual property for the sake of training these models (which tech companies like OpenAI have been accused of doing). Regarding the latter, this basis of largely stolen content should make us all the more cautious when characterizing AI-generated content as ‘new’ and, worst of all, presenting it as our own work; but anticipating that both will continue to be done, AI disclaimers should make their way into all kinds of content creation to mitigate that harm.

⁶⁸ In terms of tools born out of technological advancements, consider the electronic calculator: at its birth in the 1960s, the device was met with some of the same worries and controversies as AI today regarding the fate of human knowledge and our reliance on

the tool takes, or is given, the role of the thinking user, then it does not merely *help* do the work, it *does* the work. The role of the human user is then reduced to copy pasting, rewording or restructuring at most, as the “generative” or creative part of the work has already been taken care of. This may seem harmless when the “work” or task in question is just writing an email; but what about essays? Cover letters? Speeches? Given that its outputs are often satisfactory, reliance on AI to generate the work just once can easily evolve into an overreliance, until we find ourselves not writing anything without it anymore. By that point, no wonder even “decent writers” come to regard themselves as less capable and therefore in *need* of AI, which only further entrenches the dependence. Beyond disinterest towards the activity of thinking (considered in Part I), sophiaphobia has reached its paroxysm in the age of generative AI as those disinterested in thinking for themselves can now *delegate* the activity altogether – a possibility which more and more users are regularly seizing upon, at the detriment of both the quality of

technology. Yet a calculator can only *help* us do the calculations – it cannot solve the whole problem for us. Consider the Internet as well, which sparked similar fears as it became increasingly accessible at the turn of the twenty-first century. Yet the Internet only *helps* us access information. In contrast, generative AI can both 1) help us do calculations or find certain information *and* 2) solve the whole math problem or produce the whole essay, report, etc. that we were to produce ourselves with the help of that information.

their work and the sharpness of their own, decreasingly exercised mental capacities.

One may dismiss these concerns by pointing out the productivity gains offered by generative AI, especially when it comes to writing. I will call this the productivity argument. Indeed, why spend time writing something as formulaic as an email when it can be written for me within seconds? Why spend hours crafting this essay for some required course in which I have no interest when I can have it produced for me, and then make that text my own with some edits? If I have dozens of cover letters to write, why not give the chatbot my resume along with the job description and have it write more or less what I would have written in each case? Similarly, consider a lawyer who has been working on a case all day and is set to present on another case in court the next morning: why would they not use the generative tool at their disposal to prepare a more persuasive speech than the one they might have written themselves under such conditions? Granted, these are all valid lines of reasoning if we focus on opportunity costs: given limited time, and especially within capitalist systems wherein “time is money,” one must constantly choose between different activities based on their relative “value.” Construed in monetary terms, no “value” is directly gained from writing an email, essay or cover letter. Construed in terms of other benefits such as pleasure or knowledge, one hardly gains any “value” from writing an email or a cover letter, nor from a required essay or speech in which one has little interest. Under this framework, one could argue that generative AI actually facilitates our thinking in matters of

concern to us: allowing us to dedicate more time and brainpower to those matters by efficiently generating all kinds of other work that would have “cost” us considerable time to think through.

This argument may appeal to anyone looking to optimize their own time and accomplish those matters of lesser concern (such as formulaic emails) as efficiently as possible; in that sense, each one of us may have reasons to ask ChatGPT, Claude, and the like to do the thinking for us in certain contexts and thereby save ourselves time overall. Yet that does not mean that we should; in fact, I maintain that there are moral reasons why we should not, lest we allow thoughtlessness and sophiaphobia to be normalized on a wider scale than ever before.

First, what I find harmful behind the productivity argument, as with the comment above, is the logic of resignation behind it. Indeed, to place the task that I must complete into the “hands” of a chatbot and to then present the output as my own is to trust the chatbot’s “judgment” and writing abilities more than I trust mine; to resign myself to the supposed inferiority of my own intelligence and surrender to the superiority of *artificial* intelligence. This might seem harmless if done once, but anyone who uses generative AI or sees it used around them will know that it is never just once: as seen earlier, the line between reliance and overreliance is thin when outputs are deemed satisfactory, even more so given opportunity costs. Considering that such reliance thus tends to be repeated, resignation becomes entrenched more and more deeply each time: from ‘I don’t know what to write, let me see what ChatGPT or Claude comes up

with' to 'why spend time figuring out what to write if ChatGPT or Claude can write something better anyway.' By relying more and more on these LLMs, we gradually convince ourselves that we *need* them, when really we are perfectly capable of writing, creating, and simply thinking without them. LLMs may "explain things better than I can" in some respect (insofar as no human intelligence can compete with the breadth of an AI's "knowledge"), but that does not mean that I don't know how to explain things myself. In fact, bringing in one's unique human perspective and style will make the explanation richer and ultimately more valuable;⁶⁹ the same goes for all of the written works considered above. I would encourage each of us to hold on to this logic of self-sufficiency rather than surrendering to the logic of resignation, lest we devalue our own intelligence to the point of barely ever using it anymore.

Second, the productivity argument assumes that it is a waste of time, all things considered, to dedicate oneself to tasks and assignments that an LLM can accomplish quicker and better. Having addressed the latter issue, let us now address the former. What I find harmful here is not the idea that LLMs are quicker "thinkers," as that is factual by design: just as no human can

⁶⁹ This echoes some of the arguments for why AI-generated "art" cannot be considered real art, or why the latter can only emerge from a (human) creative *process*. The literature defending this position is plentiful and certainly worthy of attention as further grounds to reject the logic of resignation.

compete with the breadth of an AI's "knowledge," it is physically impossible for even the most quick-thinking human to "generate" quality content at the same speed as an LLM. What I do find harmful here is the positioning of efficiency as an end in itself and the justification of all means towards it. Indeed, painting time management as the only concern leads to teleologically justifying the repeated use of generative AI for *anything* that involves even the least amount of critical or creative thought: given that such thinking requires dedicated and extended time, indulging in it goes against the logic of "time is money" and its emphasis on speed for the sake of efficiency. In fact, philosophy runs counter to that logic by definition: in a means-end framework, "wisdom" is the end which philosophers seek, and the means to that end may not be the fastest nor the most efficient from a "time is money" perspective; such "means" may include spending years completing a PhD for the sake of knowledge, perhaps dedicating oneself to an underpaid teaching position, but most of all, spending incalculable amounts of time thinking problems through. If the active exercise of one's rational capacities is crucial for staying connected to our own humanity and namely to our moral compass, then philosophers are best positioned in that regard. Conversely, those who adhere to the productivity argument are worst positioned: that is, most likely to alienate themselves from their own humanity and moral compass by prioritizing "productivity gains" and thereby taking less and less time to think *for themselves*. Thus, the productivity argument amounts to a form of sophiaphobia insofar as it is based on the devaluation of *independent*, rational thought when

it comes to creating original work. The call to think *for oneself* appears all the more urgent against the risk of settling into the kind of thoughtlessness we may reach by having LLMs “think” for us wherever prolonged thinking is involved.

At this point, one may present another objection to my concerns: that I am exaggerating the scope of the issue – that most people still write the bulk of their own work themselves and that having an LLM quickly generate one’s formulaic emails is not nearly as dramatic as I have presented it in terms of “sophiaphobia” and “thoughtlessness.” With regards to that last point, I concede that AI-generated emails are not the problem; however, I maintain that they are a significant symptom of one. Indeed, if the problem is our tendency to rely on LLMs to think for us rather than thinking for ourselves, then aren’t AI-generated emails proof of how pervasive the problem has become? That one would even have the inclination to delegate email-writing to an LLM is a testament to how reliant we have become on these chatbots, their “judgment” and their capacity to do our own work for us: productivity argument aside, if a person cannot even formulate relatively short written communication without the “help” of AI, then no wonder they will struggle with formulating their thoughts in longer, more complex pieces of writing such as essays or cover letters – hence the temptation to delegate the thought-formulation in those cases as well. In that sense, AI-generated emails reveal a particularly pervasive and normalized form of “sophiaphobia” fueled by the logic of resignation: one that involves fleeing the activity of thinking before even engaging in it, which is anticipating the mental effort

ahead and preemptively delegating it to a chatbot under the presumption that it will produce a “better” result anyway.

That most people still write the bulk of their own work themselves, as my opponent retorts, is an idea that would require empirical analysis to be confirmed or disproved. Yet whatever the findings may reveal, I would still maintain my warning as I do now, no matter how “exaggerated” it may seem: that to rely on LLMs to generate our written work and communications for us is to embrace thoughtlessness both in ourselves and in others. Indeed, it isn’t just about myself as a lone user – applying Kant’s principle of universalizability to my own overreliance on AI would be to will a world in which *everyone* relies on AI to that degree: a world in which everything we write to each other would be ‘authored’ by AI; where no essay, cover letter, or speech would carry the voice of its supposed human author; where everything would be mediated by AI insofar as we would all collectively surrender to the superiority of its “intelligence” over our own. In fact, it is not impossible to imagine a future where generative AI becomes so developed and pervasive that it allows us to never have to think anymore. Even the more reticent individuals may be forced to give in lest they fall behind – a phenomenon which we already see happening today: indeed, like the commenter above, I may see myself as a “decent writer,” yet become discouraged when I receive lesser results than others for my work compared with technically flawless, *artificial* work. In this way, we may *all* be persuaded that having AI think for us is to embrace “better” work and higher productivity – which is why my warning

would remain of utmost urgency even if reliance on LLMs were still a rarity.

Thus, to those who may still think I am overstating the potential effects of generative AI, my advice would be to stay aware and critical, while keeping Arendt's insights in mind. Indeed, though her portrayal of Eichmann was factually incorrect, her conclusions drawn from his behavior should not be discarded if we take them as a 'cautionary tale' of what can happen to someone who truly does embrace thoughtlessness. Following this 'tale,' let us consider what we may learn from "Arendt's Eichmann" and from applying the "banality of evil" to AI-powered thoughtlessness today.⁷⁰ Unlike Hitler, Arendt's Eichmann did not become one of the major organizers of the Final Solution by way of explicitly evil motives. Rather, he reached that point through a gradual process of alienation, away from his own humanity and into the role of a "desk murderer"⁷¹ who went to work, did what he had been hired to do, went home, and repeated the same routine the next day — without ever stopping to think through exactly *what* he was doing and *why*. In

⁷⁰ For another way in which the "banality of evil" may be considered an enduringly relevant concept in the age of generative AI, particularly with regards to the harmful or immoral "advice" that chatbots can output when prompted, see Bess Kalb, "ChatGPT is the banality of evil," *The Grudge Report*, March 6, 2026.

⁷¹ Robert Merrihew Adams, "The Concept of Evil," *Stanford Encyclopedia of Philosophy*, last modified February 23, 2023.

psychological terms, Arendt's Eichmann began to act thoughtlessly when he stopped exercising his brain: indeed, studies show that the human brain has muscle-like qualities, such as the need to exercise it in order for it to work properly. The more a person uses their brain and challenges it with reasoning-based tasks, the better it will be at thinking.⁷² Conversely, if a person stops challenging their own brain due to laziness, 'lack of time' or a logic of resignation, and thereby lets LLMs do all the thinking for them, then their brain will gradually become weaker. This might make us more likely to mindlessly accept unethical tasks from our employer, or immoral policies and social arrangements imposed by the ruling class – until, before we know it, we fall into a routine of organizing or contributing to the deaths of faraway people from the comfort of our own desk, without ever really stopping to think about what we are doing.

As a possible safeguard against the problem of thoughtlessness and its grave moral implications, I will conclude with a call to "return to thinking itself," to repurpose the famous rally cry of the phenomenological movement; that is, a call to philosophize. As Wittgenstein once wrote, "philosophy is not a theory but an activity."⁷³ Indeed, far from being just an academic

⁷² "Is the Brain a Muscle?" *Your Heights*.

⁷³ Wittgenstein, Ludwig. *Tractatus Logico-Philosophicus*. Translated by C. K. Ogden. London: Kegan Paul, Trench, Trubner & Co., 1922, 70.

discipline or a university department, philosophy is a daily practice and an active outlook on life: it is to always examine and question the world around us, to never accept things uncritically, to never stop “talking about fundamental issues necessary for improving oneself and others”⁷⁴ as Moore put it. In that sense, to philosophize is to “*stop and think*,” to use Arendt’s formulation. Contrary to what rigid academic boundaries suggest, philosophizing is not reserved for philosophy students: on the contrary, I exhort every individual to engage in philosophical practice in some way, whether it be by reading the classics, taking a course, or cultivating that critical outlook on the world, never shying away from those ‘deep’ conversations with others and with oneself. By maintaining this critical mindset, we protect ourselves from falling into thoughtlessness and the evil-doing which it can facilitate. Similar to how Plato presented Socrates as an ideal at a time when *philosophoi* were hardly respected, I argue that we should all follow the example of philosophers – in terms of constantly exercising and holding on to our capacity for critical, independent thought – to “condition ourselves” against evil-doing, following Arendt’s insights.

Conclusion

⁷⁴ Moore, 257.

In his treatise on the nature of education relative to the nature of man, Rousseau exhorts every instructor to give their young “pupil” the space to learn and think for themselves: “Put the questions within his reach and leave them to him to resolve,” he advises. Only then, by drawing on their natural capacities to examine the world and think through problems themselves, shall the pupil become an enlightened individual – guided by reason, and free from external influences and coercion. Conversely, Rousseau warns the instructor: “If ever you substitute in his mind authority for reason, he will no longer reason. He will be nothing more than the plaything of others’ opinion.”⁷⁵ Having explored what can happen when one’s rational capacities are de-prioritized or delegated altogether, it should now be clear why the substitution of reason is indeed a danger to be avoided and actively resisted.

Based on this understanding of human beings as fundamentally capable of rational, independent thought, the aim of this paper has been to show how the de-prioritization of this fundamental capacity is harmful not just to the individual, but possibly to society as a whole. What may begin as one’s own disinterest in taking the time or making the effort to think critically, may evolve into the delegation of that capacity whenever another “thinking” authority is at one’s disposal; in

⁷⁵ Jean-Jacques Rousseau, *Emile or On Education*, trans. A. Bloom (Basic Books, 1979), 168.

the meantime, such attitude towards the value of thinking as an activity may manifest itself as scorn or even animosity towards those who dedicate themselves to said activity above all else. Hence the connection which I have drawn between the devaluation of critical thinking and the devaluation of philosophers and philosophy as a whole, all gathered under the term “sophiaphobia.” To show the concrete manifestations and consequences of sophiaphobic attitudes towards both thinkers and thinkers itself, I have analyzed such attitudes in three distinct contexts similar in one aspect: the wide-scale or high-level embrace of “thoughtlessness” on the basis of a certain form of sophiaphobia. Through this historical analysis, I have attempted to trace the roots of the problem of thoughtlessness by returning to the origins of sophiaphobia in the West, which I have identified as the birth of the activity and discipline to which sophiaphobia is fundamentally opposed. I then examined the moral weight of thoughtlessness in light of Arendt’s analysis, in order to present it as a problem not just for self-proclaimed philosophers, but for all individuals who may either fall into thoughtlessness or be caught in the potentially destructive consequences of systems built on the banality or banalization of evil. Building on Arendt’s cautionary account, I then argued that thoughtlessness is manifesting itself more and more today with our increasing reliance on artificial intelligence rather than on our own mental capacities – a problem which I maintained as such against objections such as the productivity argument. Having previously established thoughtlessness as a fundamentally moral problem, I ultimately aimed to show why

it is crucial for each one of us to actively resist falling into such “absence of thinking” – lest we risk losing ourselves and our humanity within evil structures that no sound moral judgment based on critical thinking could ever justify.

Against this risk of alienation from ourselves as rational, moral agents, and considering the ever-evolving nature of generative AI, the call to think for oneself is as urgent as ever. Insofar as to think for oneself is to embrace the kind of critical, independent thinking to which philosophers dedicate themselves, this call can also be understood as a call to philosophize – a resounding *Sapere Aude!* or “Have the courage to use your own reason!” as Kant articulates it in his essay “What is Enlightenment?” (1784) Indeed, within a societal context where sophiaphobia is politically and socially upheld – from the scorn and hostility towards the *philosophoi* in Ancient Greece to the widespread embrace of generative AI today – to use one’s own reason and resist its substitution for another “authority” is, indeed, a form of “courage” – especially considering how, both then and now, *philosophoi* and philosophically-inclined people altogether constitute a minority. To be “enlightened,” in such contexts, is to recognize the value and moral urgency of exercising one’s own rational capacities, as time-consuming and mentally strenuous of an effort as it may be. Conversely, if we tell ourselves that we do not have the time, the energy or even the capacity to “stop and think” for ourselves, then we gradually lose the “inclination” to do so in all areas of life. Before even thinking a thought through, it becomes increasingly tempting, almost automatic, to resort to the “authority” of LLMs – thus

devaluing our own intrinsic “faculty of judging” in favor of these chatbots whose judgment is just as artificial as their overall “intelligence.” From there, if we cannot even think a thought through without turning to them, then no wonder we resort to these chatbots so automatically, almost like an extension of ourselves, when we have essays or speeches, even emails, that is *anything* moderately complex to write. Yet if we fall into such a habit, then *when* will we ever think or reason independently anymore? When will we exercise that capacity which keeps us distinctly human, cognitively alive, and morally sound? To avoid falling into this level of thoughtlessness (which each one of us can fall into, even those more philosophically-inclined, in the age of generative AI), I advance the call to philosophize as a moral precept to which we should *all* commit – not just within a class, a term paper or one particularly ‘deep’ conversation under the influence, but as a lifelong *modus operandi* – for the sake of living an “examined” life while collectively forming a healthy, reflective, more human society.⁷⁶

⁷⁶ Though it has been significantly revised for publication in *The Oracle*, whose editorial board I sincerely thank for their guidance and patience, this paper would perhaps never have existed at all without the open-ended term paper assignment which sparked my crafting of this topic in Winter 2025. Having originally written this paper for that course, I would like to thank its professor, Dr. Oran Magal, for the creative freedom which he encouraged through that assignment and for his guidance during the brainstorming process: in particular, for having introduced me to Moore’s book referenced throughout Part I, and for his help in structuring my ideas around thoughtlessness into Parts II and III

References

- Arendt, Hannah. *Eichmann in Jerusalem: A Report on the Banality of Evil*. New York: Penguin Classics, 2006. First published 1963 by The Viking Press.
- Arendt, Hannah. *The Life of the Mind*. New York: Harcourt Brace Jovanovich, 1981. <https://maytemunoz.net/wp-content/uploads/2016/10/hannah-arendt-the-life-of-the-mind.pdf>.
- Arendt, Hannah. "Thinking and Moral Considerations: A Lecture." *Social Research* 38, no. 3 (1971): 417–446. https://cooperative-individualism.org/arendt-hannah_thinking-and-moral-considerations-1971-autumn.pdf.
- Calder, Todd. "The Concept of Evil." *The Stanford Encyclopedia of Philosophy*. Winter 2022 Edition. Edited by Edward N. Zalta and Uri Nodelman. <https://plato.stanford.edu/archives/win2022/entries/concept-evil/>.
- Campbell, Martha. "Why Is He Thinking? Did He Lose His Smartphone?" *CartoonStock*, March 21, 2017. <https://www.cartoonstock.com/cartoon?searchID=CS203172>.
-

based on historical chronology. I am also grateful to Dr. Magal for his helpful comments on the first draft of this paper, as well as for his lasting contribution to the development of my philosophical thinking and writing.

- Chakrabarti, Meghna. "The Eichmann Tapes and the Comforting Myth of the 'Banality of Evil.'" *On Point* (WBUR), July 15, 2022.
<https://www.wbur.org/onpoint/2022/07/15/the-eichmann-tapes-and-the-nature-of-evil>.
- "Deaths of Philosophers." *Wikipedia*. Last modified January 2026.
https://en.wikipedia.org/wiki/Deaths_of_philosophers.
- "The Influence of Kantian Philosophy on Today's World." *Philosophy Nest*, November 18, 2025.
<https://www.philosophynest.com/foundational-philosophical-concepts/the-influence-of-kantian-philosophy-on-todays-world.html>.
- "Is the Brain a Muscle?" *Your Heights*. Accessed January 31, 2026. <https://www.yourheights.com/blogs/health/is-the-brain-a-muscle>.
- Kalb, Bess. "ChatGPT is the banality of evil." *The Grudge Report*, March 6, 2026.
<https://besskalb.substack.com/p/chatgpt-is-the-banality-of-evil-and>
- MacAskill, William. "EA: A View from the Inside." *Effective Altruism Forum*, November 17, 2020.
<https://forum.effectivealtruism.org/posts/sanhCrJohGjAyAxLr/ea-a-view-from>.
- Moore, Christopher. *Calling Philosophers Names: On the Origin of a Discipline*. Princeton, NJ: Princeton University Press, 2020.
- Rousseau, Jean-Jacques. *Emile or On Education*. Translated by Allan Bloom. New York: Basic Books, 1979. First published 1762. <http://naturalisms.org/phil-editions/modern%20to%201950/France/Rousseau%201979%20Emile,%20or%20On%20Education.pdf>

Schiff, Jade Larissa. "Thoughtlessness." In *Burdens of Political Responsibility: Narrative and the Cultivation of Responsiveness*, 50–84. Cambridge: Cambridge University Press, 2014.

<https://www.cambridge.org/core/books/burdens-of-political-responsibility/thoughtlessness/B96F588A2C089D57E35E1DC339814F34>.

"The Influence of Kantian Philosophy on Today's World." Philosophy Nest, November 18, 2025.

<https://www.philosophynest.com/foundational-philosophical-concepts/the-influence-of-kantian-philosophy-on-todays-world.html>.

Tömmel, Tatjana, and Maurizio Passerin d'Entrèves. "Hannah Arendt." *The Stanford Encyclopedia of Philosophy*. Spring 2025 Edition. Edited by Edward N. Zalta and Uri Nodelman.

<https://plato.stanford.edu/archives/spr2025/entries/arendt/>.

Wittgenstein, Ludwig. *Tractatus Logico-Philosophicus*. Translated by C. K. Ogden. London: Kegan Paul, Trench, Trubner & Co., 1922. <https://standardebooks.org/ebooks/ludwig-wittgenstein/tractatus-logico-philosophicus/c-k-ogden>.

Marx and Engels on Nature: Through Darwin, against Malthus

References to nature and the natural world are frequently made throughout Marx and Engel's body of work, however, it is seldom analyzed or commented on by readers of the history of philosophy. Furthermore, the most popular works that do treat this facet of the two theorists often glance over their theoretical engagement with Charles Darwin. Commentators such as Edward Aveling, Terrence Ball, and Ralph Colp Jr. have all proposed that the three thinkers simply exchanged a few pleasantries through a small number of letters and are only mistakenly and artificially posited to have taken philosophical ideas from each other. I argue that whilst it is true that Darwin had little to no interest in the two communists, that Marx and Engels both took substantial material from the evolutionist when it comes to their understanding of nature and the role that it plays within the large of their system. When reading through the first volume of Das Kapital, Darwin's impact is obvious, and when reading Engel's later works such as The Dialectic of Nature it becomes glaring and impossible to dismiss. In short, I draw out how the mature Marx and Engels understood nature through a Darwin stripped of his Malthusian assumptions.

Charles Darwin drew heavily from Thomas Malthus's economics to develop his theory of evolution. However, what is

to happen to Darwinian evolution and, more importantly, the conception of nature, when foundational Malthusian assumptions about scarcity, production, and subsistence are discarded and rejected? Economics, primarily through the works of Ricardo, Say, Mill, and Marshall, quickly rejected Malthus's population theory, as it asserted that production is entirely bound by factor inputs, without accounting for exogenous technological advancements—such as fertilizers and pesticides. In plainer language, Malthus's notion that population growth and economic well-being depend on competition over a fixed food supply was plain wrong. However, competition over scarce and fixed resources is one of the three pillars of Darwin's theory, but does this make evolution an untenable position? Marx and Engels were amongst the first to spot this failure in Darwinian logic, but rather than dismissing the entire system, they surprisingly crucially relied on it to understand nature. Their conception of nature shifted after the publication of *On the Origin of Species*, which had important implications for their critique of political economy and their defense of dialectics. I argue that the later Marx and Engels understood nature through Darwinian evolution, but against Malthusian population theory, and that this partly shaped their philosophically revolutionary analysis of socio-economic relations. I contrast Marx's understanding of nature, pre-Darwin, to how it appears post-Darwin in *Capital* and other later works, and finally to how this understanding is presented in Engels's later dialecticalization of nature. Parallels will also be drawn in passing between some of the three thinkers'

shared positions against essentialism and on the historicity of change.

In *The Origin*, Darwin builds what he found to be three necessary conditions for evolution to take place in nature, which, for him, explains the existing multiplicity and origin of species without appealing to special creation – the view that species are well defined and independently created by God. Firstly the condition of **variation**, that is, difference in physiological traits between common species members and between parents and offspring, which is an evident empirical fact, both in domestic and wild organisms. Importantly, as well, variances in domestically bred organisms only occur spontaneously, rather than breeders producing them themselves, and thus, variance can be attributed solely to blind natural processes¹. Second, **inheritability**: Darwin makes the case for inheritability by first showing that the inheritance of traits is true of human breeding. He uses different breeds of domesticated pigeons as an example and gives an overview of the many types of breeds and how certain characteristics, such as colour, beak form and shape, esophagus size, and others², are all passed on generationally, whether it be the English carriers elongated eyelids, the runt pigeon's size and large feet, the barb's short and wide beak, the

¹ Darwin, *On The Origin of Species*, p.44.

² Ibid. pp. 24-25.

list goes on.³ Darwin argues that the facts that pigeon breeds are interfertile (different breeds can procreate between themselves and produce healthy fertile offspring), share basic bone structure, and exhibit signs of reversion to a common form when interbred, all point to the plausibility that all breeds are descended from a common ancestor, namely the rock pigeon (*Columba livia*).⁴ This leads to the final Darwinian requirement for the evolution of species, **natural selection**. The role of natural selection in the theory is to explain why certain variable traits are kept and passed down, rather than discarded in the next generation. In other words, there must be an operative mechanism in nature which mimics the human breeder's manual [conscious] selection for Darwin to complete his analogy that there is a process of evolution in nature. This is where Malthus's theory of population comes in.

Malthus understood the economy and agricultural resources as being limited and fixed and geometric in their growth. This led him to be an advocate for tariffs, to keep exports above imports,⁵ and to the abolition of the "poor laws", which he thought bettered the poor's conditions and thus increased childbearing without an equivalent increase in the food supply

³ Ibid. p. 24.

⁴ Ibid. pp. 30-31.

⁵ Thomas Malthus, *The grounds of an opinion on the policy restricting the importation of foreign corn*. p. 42.

which hurts everyone in the long run.⁶ Malthus, like most of his 19th-century academic peers, wanted to uncover the laws that govern events, and in his view, the laws that dictate the population and thus the conditions of humankind. He postulated two principal notions: 1) The “passion between the sexes⁷ will never change, hence childbirth will continue to grow exponentially, since two people can have more than two children⁸ and 2) Food supply can only grow linearly,⁹ since only roughly the same amount of agricultural produce can be yielded from each additional acre of farmland.¹⁰ This led to Malthus’s best-known and arguably most important tenet: if a population is left unchecked, it will outgrow its means of subsistence, and there will necessarily be a surplus population [relative to food levels] which will perish and/or struggle immensely.¹¹ The latter is what Darwin appropriated into his system as the operative mechanism to explain how particular trait variances are preserved in the evolutionary process of species. More explicitly, in the necessary competition between organisms for limited amounts of subsistence goods (the struggle for existence),

⁶ Thomas Malthus, *An Essay on the Principle of Population*. p. 26.

⁷ Ibid. p. 4.

⁸ Ibid.

⁹ Ibid. p. 8.

¹⁰ Ibid. p. 7.

¹¹ Ibid. chap. 18.

features will be naturally selected and passed on based on the advantage that they grant the organism in the competitive struggle for existence.¹²

As Darwin admits in *The Origin*, his conception of competition in nature “is the doctrine of Malthus applied with manifold force to the whole animal and vegetable kingdoms”.¹³

A few important notes must first be made in passing about Darwin's resulting ideas on nature before jumping over to Marx and Engels. An important ontological move in Darwin¹⁴ was the rejection of essentialism,¹⁵ which is the notion that species have unchanging and determined essences. The dominant scientific discourse in Darwin's time understood species as being specially created, that is species-types do not change and are rather created whole and are preserved as such throughout time.¹⁶ That is not to say that special creationists denied that particular organisms change whatsoever. Rather, they asserted that species

¹² Darwin, *On The Origin of Species*, chap. III.

¹³ *Ibid.* p. 79.

¹⁴ Whilst Darwin cannot be said to have toppled bio-essentialism completely and definitively on his own, he certainly had an important historical role in it. Refer for example to Ernst Mayr's 1982 *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*, chap. 9.

¹⁵ The term “essentialism” and its variations was developed in the 20th century but has been frequently retroactively read into past philosophy. Refer to Mary Winsor's 2006 *The Creation of the Essentialism Story: An Exercise in Metahistory*, amongst other works.

¹⁶ Refer to figures such as Richard Owens, Adam Sedgwick, and Georges Cuvier.

are fixed categories, and thus particular species-type cannot change to another specie type, no matter the timeframe. Darwin made the crucial move to dislodge species from having a fixed ontological identity and rather to posit that they are fluid and changing gradually over time, whilst never being completely fixed as a type. He even went so far as to claim that their semantic referring to as distinct “species” is not scientifically grounded:

From these remarks it will be seen that I look at **the term species as one arbitrarily given**, for the sake of convenience, to a set of individuals closely resembling each other, and that it **does not essentially differ** from the term variety, which is given to less distinct and more fluctuating forms.¹⁷

Another important facet in Darwin, which will be important to Marx and Engels, is the importance of the historicity of change. Darwin injected into the scientific method the contingency of history and how an organism’s particular material position on the historical plane shapes its development by a unitary process. This, again, was against the prominent scientific ideology of the era, which took development as being teleological; things, both animate and inanimate, change not because of their immediate conditions and afflictions, but

¹⁷ Darwin, *On The Origin of Species*, p.66.

because they are progressing towards an endpoint.¹⁸ The above is what Marx and Engels will principally draw on in their engagement with Darwin and nature.

The Darwinian line that cuts across Engels and Marx is often implied to simply be a co-inhabitancy in the 19th-century academic scene, and that there was no serious theoretical engagement with each other, rather only a few exchanged pleasantries through written addresses.¹⁹ Whilst Darwin indeed had no philosophical interest in Marx and Engels's works,²⁰ as written earlier, this essay holds that a closer literary examination of Marx and Engels shows that Darwin did certainly influence their position on nature, even if it isn't always explicitly stated. The same is true of Engels; however, his connection to Darwin is more explicit in his posthumous work *Dialectics of Nature*. For the sake of historical context, Marx and Engels started publishing and developing their ideas in the mid-1840s and started working on the first volume of *Capital* in the early 1860s, and Darwin

¹⁸ Refer to thinkers such as William Payley and Robert Chambers, who helped develop notions which led to the view that the physical universe is a contrivance that is teleologically oriented towards a higher formal organization. Particularly Payley's 1802 *Natural Theology*, and Chambers's 1844 *Vestiges of the Natural History of Creation*.

¹⁹ See for example Terence Ball's 1979 *Marx and Darwin: A Reconsideration*, Edward Aveling's 1897, *Charles Darwin and Karl Marx: A Comparison*, and Ralph Colp Jr.'s 1974, *The Contacts Between Karl Marx and Charles Darwin*.

²⁰ See Darwin's only letter to Marx, written October 1st, 1873, after receiving a copy of the first volume of *Capital*.

published his major work *On the Origins of Species* in 1859. Engels reads *The Origin* shortly after its publication and writes to Marx: “Darwin, by the way, whom I'm reading just now, is absolutely splendid. There was one aspect of teleology that had yet to be demolished, and that has now been done”.²¹ Marx, within the same year, writes to Lasalle, “Darwin’s book is very important and serves me as a basis in natural science for the class struggle in history”.²²²³ These latter letters were amongst the first of 38 [known] letters written by either Marx or Engels that mention Darwin.²⁴ Common themes found in many of these letters are 1) they appreciate and take Darwin’s rejection of teleological progress in nature, 2) they like Darwin but reject his adoption of Malthusian notions of population and competition (since they understood competition of this sort to belong to capital rather than nature), and 3) natural development and economic modalities parallel each other in their being historically contingent. The rest of this essay will thus dig out and explain the above themes in various Marxian works that came before *The*

²¹ Pitskhelauri, M. (Ed.). (2010). *Marx-Engels Collected Works* (Vol. 40). Lawrence & Wishart. p. 550.

²² Rudenko, N., Vladimirova, A. (Ed.). (2010). *Marx-Engels Collected Works* (Vol. 41). Lawrence & Wishart. p. 245.

²³ Marx makes very similar remarks on Darwin in multiple letters. A second appearance of that sentence appears both in a letter to Lasalle and in another to Engels.

²⁴ Bergman, Jerry, *Friedrich Engels Introduced Darwin to Karl Marx and Changed the World*, p. 463.

Origin and in works that came after, to show through contrast how Darwin did have a concrete influence on Marx and Engels's understanding of nature.

A consistent conception of nature is found in earlier Marx works, which differs from his and Engels' later texts (post-*Origin*). In the first manuscript of the *Economic and Philosophic Manuscripts of 1844*, Marx treats nature and science extensively and materialistically; "Nature is man's inorganic body – nature, that is, insofar as it is not itself human body".²⁵ Here, Marx draws a sharp line between "nature" and human. Nature, at this point, for Marx, is a stable, homogenous background over which human activity and production dances. (Incidentally, Engels held a very similar position in his early texts, such as in his 1845 *The Condition of the Working Class in England*.) That is not to say that humans and their bodies fall outside of nature; human is of course, completely encompassed and situated by/in nature, but what makes humankind such a distinctive species is precisely the fact that she can intake any organic material into her sphere of activity, a concept Marx calls universality.²⁶ In simpler terms, animals, as opposed to humans, can only interact with nature in an immediate instinctual way, whereas for humans, everything they encounter in nature can be coerced into their domain through labor, whether it be science, art, or anything else

²⁵ Marx, *Economic and Philosophic Manuscripts of 1844*, manuscript 1, p. 31.

²⁶ *Ibid.* p. 31.

conscious. Furthermore, a few pages later, Marx writes that “through production, nature appears as [humans’] work and his reality. The object of labour is, therefore, the objectification of man’s *species-life*: for he *duplicates himself* not only, as in consciousness, intellectually, but also actively, in reality, and therefore he sees himself in a world that he has created”.²⁷ (An important underlying notion here, that will not be discussed explicitly as it falls outside the scope of this essay, is of course that human consciousness is materially engaged and generated through a dialectical interplay).²⁸ The meat of Marx’s early conception of nature can thus be summed as follows: 1) Nature in the human sphere is praxiological and anthropocentric, that is, it serves only as the stage for human activity whilst not being active itself, and 2) Human, as an organism, is a species-type, who’s essence is defined by it’s “free [and] conscious activity”,²⁹ i.e., labor.³⁰

Marx continues to mention nature in various works after the 1844 manuscripts but before the publication of *The Origin*.

²⁷ Ibid. p. 32.

²⁸ Human is thrown into the sensuous, which then is redoubled onto and into nature (through consciousness) which quilts Nature. Refer to *ibid*, p.48.

²⁹ Ibid. p. 31.

³⁰ Note: It is clear here what alienation, even in the early Marx, refers to; The inhibiting of the material expression of labor and/or the restriction of freedom and/or consciousness in production.

During this “middle” period, Marx did admittedly rework the above ideas, presumably due to his [even] further distancing from Feuerbach and Hegel during this time. For example, just a few years later in *The German Ideology*, he overturns his previous essentialist conception of humankind:

Men can be distinguished from animals by consciousness, by religion or anything else you like. They themselves begin to distinguish themselves from animals as soon as they begin to produce their means of subsistence, a step which is conditioned by their physical organisation. [...] *The nature of individuals thus depends on the material conditions determining their production.*³¹

Essence is thus annulled and instead of Human being a species-type, its subjective structuring is instead crafted as being the relation between the individual and its historically contingent production mode. However, whilst the big of the notion changed, it did not change nature itself in Marx; it is still presented as being a static resource plane.

The major change in Marx’s conception of nature itself comes only after Darwin’s major publication, principally in the first volume of *Capital*. In a well-known footnote of the fifteenth chapter of the book, Marx writes:

³¹ Marx, 1846, *The German Ideology*, section I, p.3.

*Darwin has interested us in the history of Nature's Technology, i.e., in the formation of the organs of plants and animals, which organs serve as instruments of production for sustaining life. Does not the history of the productive organs of man, of organs that are the material basis of all social organisation, deserve equal attention? [...] Technology discloses man's mode of dealing with Nature, the process of production by which he sustains his life, and thereby also lays bare the mode of formation of his social relations, and of the mental conceptions that flow from them. [...] The weak points in the abstract materialism of natural science, a materialism that excludes history and its process, are at once evident from the abstract and ideological conceptions of its spokesmen.*³²

This unassuming footnote can be treated as a summary of the whole of Marx's³³ latest conception of nature. It explains how it relates to Darwin, how it goes against Malthus, and what space it occupies in his larger system. As the footnote shows, there are now three distinct onto-logical spaces in the world for Marx; the

³² Marx, 1867, *Capital* vol. I, chapter 15, footnote 4.

³³ And presumably Engel's at the time. However, due to the homogeneity of their writings and contributions in *Capital*, it becomes often impossible to make claims about clear differences in their positions at the time.

first being, of course, purely human social activity, the second is purely natural change, and the third is where the two spheres are united through production. As Alfred Schmidt puts it, “natural and human history together constitute for Marx a differentiated unity. [...] human history is not merged in pure natural history; natural history is not merged in human history”.³⁴ This is a complete turning away from the Marx in the 1844 manuscripts, where nature was inorganic and only turned active through human praxis. Darwin thus clearly influenced Marx to incorporate an account of nature and natural processes which falls completely outside of human engagement, a massive ontological reconfiguring in the large of the theory.

Marx and Engels’s vehement rejection of Malthus is also of importance in this new understanding of the natural world. They dedicated hundreds of pages throughout their works to dismantling Malthus on an economic basis, and they subsequently heavily criticized both Darwin himself and other related thinkers³⁵ for making the Malthusian move of reading competition into nature. In an 1862 letter to Engels (in the early stages of writing *Capital*), Marx writes:

Darwin rediscovers, among the beasts and plants, the society of England with its division of labour, competition,

³⁴ Schmidt, *The concept of Nature in Marx*, p. 45.

³⁵ E.g. Herbert Spencer, Thomas Huxley, and William Sumner, amongst others.

opening up of new markets, 'inventions' and Malthusian 'struggle for existence'. It is Hobbes' *bellum omnium contra omnes* and is reminiscent of Hegel's *Phenomenology*, in which civil society figures as an 'intellectual animal kingdom', whereas, in Darwin, the animal kingdom figures as civil society.³⁶

Marx here rebukes Darwin's assertion of the struggle for existence, as formulated by Malthus, as being universal natural law since he deems it to be a particular feature of capitalist logic. However, Marx did still admit that there does seem to be a population law in action in the natural world, since, as shown above, he now places "pure" nature as also having its own logical space; "in fact every particular historical mode of production has its own special laws of population, which are historically valid within that particular sphere. An abstract law of population exists only for plants and animals".³⁷ In other words, Marx took up Darwin's theory to fuel his conception of nature, but he does so by first rejecting the Malthusian elements in the theory, which leaves him with the understanding of nature that was fleshed out earlier in this paragraph. Nature, for the mature Marx, thus did operate and develop non-teleologically, nonessentially, and

³⁶ Rudenko, N., Vladimirova, A. (Ed.). (2010). *Marx-Engels Collected Works* (Vol. 41). Lawrence & Wishart. p. 380.

³⁷ Marx, *Capital volume I*, (penguin translation), p. 784.

independently from human consciousness, as in Darwin. However, contrary to Darwin, nature's own internal laws and operations did not transfer over to where it touched the social (production and labor), due to humankind's rupturing from pure nature through its dialectical redoubling³⁸ (i.e. economic activity).

Engels was more engaged with Darwinian theory than Marx, both as it appeared in biology and social analysis ("social Darwinism"). He put considerable effort later in his career, post-*Capital*, to crafting his own scientific account of nature, which in the end distanced him from the conception of nature found in his joint works with Marx. He did this mainly in two steps, first in *Anti-Dühring* (1878) and finally in *Dialectic of Nature* (written in the 1880s). Both of these latter works draw from Darwin heavily and even provide interpretations and defenses of his thought.³⁹ But Engels did not completely abandon Marx and his earlier ideas; he rather stepped off of their previous notions to end up with a conception of World as being totally dialectical, rather

³⁸ Dialectical redoubling is contemporarily understood as the negation of negation, by which an object's identity is constituted retroactively through an Other. For Marx here, nature as undefined and indetermined gives rise to humans, who then through their own doings end up defining and determining what Nature is through material activity. However, Marx seems to miss that nature as held in human practices is what retroactively generates the seeming possibility of a primary "pure" nature. See Žižek's *Less Than Nothing* for more.

³⁹ E.g. *Anti-Dühring*, p. 48, and part I, chap. 7, respectively.

than being dialectical only in economic and social activity. In a fragmentary note from *Dialectic of Nature* he writes:

Dialectics, so-called *objective* dialectics, prevails throughout nature, and so-called subjective dialectics, dialectical thought, is only the reflection of the motion through opposites which asserts itself everywhere in nature, and which by the continual conflict of the opposites and their final passage into one another, or into higher forms, determines the life of nature.⁴⁰

In short, Engels wanted to provide a dialectical framework for the natural sciences in opposition to the prevailing mechanistic basis. In *The Dialectic of Nature* Engels lays out what he deems to be three principal operative dialectical mechanisms in [pure] nature:⁴¹ 1) “The law of the transformation of quantity into quality and vice versa”, 2) “The law of the interpenetration of opposites” and 3) “The law of the negation of the negation”.⁴² He comes to these conclusions by interpreting⁴³ various findings

⁴⁰ Engels, *Dialectics of Nature*, p. 211.

⁴¹ These were also developed in *Anti-Dühring* but appeared more explicitly in *Dialectic of Nature*.

⁴² Engels, *Dialectic of Nature*, p. 19.

⁴³ Often in questionable ways, it is quite obvious in these “scientific” works that Engels was still primarily a social thinker. E.g. from *Dialectic of Nature* p. 86: “This

from the fields of biology, chemistry, and physics. The actual contents of these postulates are of no importance to this essay. However, what is important is the conceptual move made here by Engels. By this point in his thought, Engels has completely inverted his and Marx's earliest conceptualizations of nature; Rather than nature being an inorganic unidimensional plane that is only turned active through human consciousness (i.e. produced through a redoubled negation) between the inorganic and the social situation,⁴⁴ Nature and human activity now share the same ontological space whose coordinates are actually determined by nature itself. Furthermore, it was actually Darwin's evolution which paved the way for Engels's final understanding of the world (nature and human) as structurally sharing the same dialectical development. As Engels put it:

"predatory economy" of animals plays an important part in the gradual transformation of species by forcing them to adapt themselves to other than the usual food, *thanks to which their blood acquires a different chemical composition* and the whole physical constitution gradually alters".

⁴⁴ Marx and Engels early understanding of nature can fairly be described as appearing through sublation [*aufhebung*] between the inorganic and subjective activity, even though they didn't use this Hegelian term explicitly in this context. E.g. "Through this production, nature appears as his work and his reality. The object of labor is, therefore, the objectification of man's species life: for he **duplicates** himself not only, as in consciousness, intellectually, but also actively, in reality, and therefore he sees himself in a world that he has created". 1844 manuscripts, manuscript I, p. 32.

Nature is the proof of dialectics, [...] Nature works dialectically and not metaphysically; [...] she does not move in the eternal oneness of a perpetually recurring circle, but goes through a real historical evolution. In this connection, Darwin must be named before all others. He dealt the metaphysical conception of Nature the heaviest blow by his proof that all organic beings, plants, animals, and man himself, are the products of a process of evolution going on through millions of years.⁴⁵

This conception of nature in the final stretch of Engels's academic career was at the foreground of much of his last works. He went further than Marx ever did when it came to the application of their dialectical materialism and believed that nature's own dialectical development was a primary condition for humankind's own historical development, a position clearly instigated by Darwin.

All in all, there is substantial evidence for proposing that Marx and Engels were both influenced by Darwin's theory of evolution. When tracing their respective ideas and conceptions of nature, it becomes clear that Darwin is what allowed both thinkers to understand nature as they did in their mature works. *On The Origin of Species* gave Marx the foundation for positing an

⁴⁵ Engels, 1880, *Socialism Utopian and Scientific*, p. 56.

active and developing nature outside of human activity, and for Engels to place human wholly inside of nature. As the case was made in this essay, the two thinkers thus had a real and important philosophical engagement with Darwinian thought, rather than a simple symbolic mutual respect for him as it has been popularly presented. However, they also did not appropriate the whole of the theory into their own, since they still placed scarcity and competition as being part of a bourgeois logic; they thus understood nature largely through Darwin, and against Malthus.

References

- Aveling, Edward B. *Charles Darwin and Karl Marx: A comparison*. London: Twentieth Century Press, 1897.
- Ball, Terence. *Marx and Darwin: A Reconsideration*. Political Theory, Vol. 7, No. 4, pp. 469- 483. Sage Publications, 1979.
- Bergman, Jerry. *Friedrich Engels Introduced Darwin to Karl Marx and Changed the World*. Answers Research Journal 14, 463–472, 2021. https://assets.answersresearchjournal.org/doc/v14/engels_darwin_marx.pdf.
- Chambers, Robert. *Vestiges of the Natural History of Creation*. The Project Gutenberg eBook of Vestiges of the Natural History of Creation, December 1, 2004. <https://www.gutenberg.org/cache/epub/7116/pg7116.txt>.
- Colp, Ralph. *The Contacts Between Karl Marx and Charles Darwin*. Journal of the History of Ideas , Apr. - Jun., 1974, Vol. 35, No. 2, pp. 329-338. University of Pennsylvania Press, 1974.
- Darwin, Charles. *On the Origin of Species*. The Project Gutenberg eBook of On the Origin of Species, by Charles Darwin. <https://www.gutenberg.org/files/1228/1228-h/1228-h.htm#chap00>.
- Darwin, Charles. Letter to Karl Marx. "To Karl Marx 1 October 1873." Beckenham, Kent: Down, October 1, 1873.
- Engels, Friedrich. *Herr Eugen Dühring's Revolution in Science* . Translated by Emile Burns. Moscow: Progress Publishers, 1947. https://www.marxists.org/archive/marx/works/download/pdf/anti_duhring.pdf.
- Engels, Friedrich. *Dialectics of Nature*. Transcribed by Sally Ryan. Marxist internet archives. 1998.

<https://www.marxists.org/archive/marx/works/download/pdf/dialectics-nature.pdf>

- Engels, Friedrich. *Dialectics of Nature: Notes and Fragments*. Translated by Clemens Dutt. Moscow: Progress publishers, 1934. <https://www.marxists.org/archive/marx/works/1883/don/ch07c.htm>
- Engels, Friedrich. *Socialism: Utopian and Scientific*. Marx/Engels Selected Works, Volume 3, p. 95-151. Translated by Edward Aveling. Moscow: Progress Publishers, 1970. https://www.marxists.org/archive/marx/works/download/Engels_Socialism_Utopian_and_Scientific.pdf.
- Malthus, Thomas R. *The grounds of an opinion on the policy restricting the importation of foreign corn*. London: J. Johnson and co., 1815.
- Malthus, Thomas R. *An Essay on the Principle of Population, as it Affects the Future Improvement of Society with Remarks on the Speculations of Mr. Godwin, M. Condorcet, and Other Writers*. London: J. Johnson, 1798.
- Marx, Karl, and Friedrich Engels. *Marx & Engels Collected Works*. Vol. 40. London: Lawrence & Wishart, 2015.
- Marx, Karl, and Friedrich Engels. *Marx & Engels Collected Works*. Vol. 41. London: Lawrence & Wishart, 2015.
- Marx, Karl. *Economic and philosophic manuscripts of 1844*. Translated by Martin Milligan. Moscow: Progress Publishers, 1959. <https://www.marxists.org/archive/marx/works/download/pdf/Economic-Philosophic-Manuscripts-1844.pdf>.
- Marx, Karl. *A Critique of The German Ideology*. Transcribed by Tim Delaney. Moscow: Progress Publishers, 1968. https://www.marxists.org/archive/marx/works/download/Marx_The_German_Ideology.pdf.

- Marx, Karl. *Capital A Critique of Political Economy Volume I Book One: The Process of Production of Capital*. Translated by Samuel Moore. Moscow: Progress Publishers, 1867.
<https://www.marxists.org/archive/marx/works/download/pdf/Capital-Volume-I.pdf>.
- Marx, Karl. *Capital Volume I*. Vol. 1. 3 vols. Great Britain: Penguin Books Ltd, 1976.
- Mayr, Ernst. *The growth of biological thought: Diversity, evolution, and inheritance*. Cambridge, Massachusetts: Belknap Press, 1982.
- Payley, William. "Natural Theology or Evidences of the Existence and Attributes of the Deity." Early Modern Texts, March 2018. <https://www.earlymoderntexts.com/assets/pdfs/paley1802.pdf>.
- Schmidt, Alfred, and Ben Fowkes. *The concept of nature in Marx*. London: NLB, 1971.
- Winsor, Mary P. *The creation of the essentialism story: An exercise in metahistory*. Firenze: Olschki, 2006

Essentialist (Mis)givings

This paper critiques Eleonore Neufeld's essentialist account of slurs, according to which they function as failed natural kind terms that encode an imagined group essence responsible for negative stereotypical features. Slurs, on this view, are informationally rich despite having empty extensions, since they presuppose a causal “mini-theory” linking imagined essence to derogatory surface features. After reconstructing Neufeld's framework within the Kripke-Putnam view on natural kind terms, I show her theory successfully explains a wide range of canonical and non-canonical slur uses. However, it also underestimates the instability of slurs in natural language. I introduce slurs as floating signifiers by demonstrating two cases where slurs' target groups are unconventional. To accommodate these cases, I propose a modest modification to Neufeld's definition of slurs that foregrounds the speaker's perspective and agency. This adjustment preserves the strengths of essentialism while offering a more politically responsible account of how slurs are discursively weaponized.

In this paper¹, I will read Eleonore Neufeld’s essentialist account of slurs, which treats them as failed natural kind terms.² Neufeld’s account is impressive both in theoretical and interdisciplinary scope. Her work marks a shift from how the analytic tradition has traditionally treated slurs — as paradigmatic hard cases to the Fregean tradition. But it is only the start, not the end of work to be done.

Neufeld assumes the Kripke-Putnam view on natural kind terms.³ I will explain key parts of this view that will be relevant. Natural kinds are categories of nouns, referred to collectively. These categories are in turn things that exist in the actual world. UNICORN⁴ is not a natural kind, because there are no unicorns in the actual world. But HORSE is a natural kind. So is DONKEY. Each kind, which is a set consisting of multiple distinct objects, is also distinct from other kinds. The kind HORSE can be expressed using a set which lists all the horses in the world.

¹ This paper would have been a non-starter without Andriy Bilenkyy. To him I owe my most heartfelt acknowledgement.

² Eleonore Neufeld, “An Essentialist Theory of the Meaning of Slurs,” *Philosophers’ Imprint* 19. no. 35 (2019): 1–29, <http://hdl.handle.net/2027/spo.3521354.0019.035>.

³ Saul Kripke, *Naming and Necessity* (Basil Blackwell: 1980), 116–55; Hilary Putnam, “The Meaning of ‘Meaning,’” *Minnesota Studies in the Philosophy of Science* 7 (1975): 131–93, <https://hdl.handle.net/11299/185225>.

⁴ A brief note on convention: I use all capital letters (e.g. HUMAN) to designate a natural kind, single quotes (‘ ’) to designate slur terms (mention as opposed to use), and double quotes (“ ”) for standard quotations.

Let us consider the natural kind HUMAN. Had I decided to list the names of all humankind, eight billion strong in a set {Neufeld, Kripke, Putnam, Camp, ...}, I would be writing out HUMAN in extension. But if I do not have nearly enough time, I may also choose to specify the set in intension. Doing so means that I specify some property which x has iff x is a human, and there are various ways to do this. I could write $\{x \mid x \text{ is a member of } \textit{Homo sapiens}\}$.

According to the Kripke-Putnam view, natural kinds are established by fixing an exemplar and establishing similarity. To bring about HUMAN as a natural kind, I first need to find something that can serve as the archetypal representation of HUMAN. Say I pick my friend, Harold. But what counts as being similar to him? He studies physics and mathematics, but I do not. He is Canadian, but I am not. Academic interests and citizenships are not essential for humanity, then. It cannot be that the essence of HUMAN consists of socially variable factors. We want to appeal to something more innate, maybe a biological characteristic. Harold and I both have 46 chromosomes. But my other friend Gerry, who has Down syndrome, does not. So it cannot be that either.

This example is meant to illustrate that the question of essences is an open one. Even though I have succeeded in defining an intension for HUMAN as $\{x \mid x \text{ is a member of } \textit{Homo sapiens}\}$, I have not succeeded in defining what makes someone a member of *Homo sapiens*. The latter question is the question of essence. It may just as well be that the answer is a circular

definition, that $\{x \mid x \text{ bears the human essence, to be determined}\}$. This resort exemplifies the placeholder solution.

Most contemporary theorists working with the Kripke-Putnam view have adopted the stance that natural kind essences are hidden or unobservable.⁵ Neufeld herself proposes thinking of essences as to be determined, a placeholder for the time being. For me and Neufeld, slurs are perplexing in how they seem to behave as natural kinds in some ways, but not in others. Slurs can and do have clear, identifiable initial events wherein an exemplar is selected, for one (fig. 1).

⁵ As for why the essence of a natural kind term is so elusive, one answer suggests the metaphysical premiss that is presented in terms of the “qua problem.” (see Sören Häggqvist and Åsa Wikforss, “Natural Kinds and NKTs: Myth and Reality.” *The British Journal for the Philosophy of Science* 69, no. 4 (2018): 915, <https://doi.org/10.1093/bjps/axw041>.) Every speaker has a distinct psychological state (“the mental state of the speaker”) which, in concert with their community of practice, determines what their exemplar is and who to defer to. This social externalism of the speaker determines the semantic externalism of the natural kind term in question. If proper names obtain in virtue of an initial baptism event, and then through a causal chain where a speaker intends its use to be identical to the previous speaker in a perennial state of ostension and deferral *à la* Kripke, then it is clear no single causal chain can exist for natural kind terms.



Figure 1. An internet meme illustrating how an exemplar may be selected.⁶

Though she is rightfully proud that her view “does not stand or fall on the basis of a single piece of evidence”, I will show how her account can be strengthened by considering slurs as floating signifiers.⁷ I define floating signifiers as nouns (including nominalizations) which refer not to multiple, but an indeterminate number of referents.

Floating signifiers are special not only due to polysemy. Neufeld attributes to all slurs *S* negative evaluative properties. *S* is said to encode a set of derogatory features embedded semantically or conventionally in itself, the lexical item. I argue that the current framework relying on these features is

⁶ Prestigious_Cat2052 (u/Prestigious_Cat2052), “me_irl,” Reddit, November 30, 2025, https://www.reddit.com/r/me_irl/comments/1paoz81/me_irl/.

⁷ Neufeld, “An Essentialist Theory,” p. 23.

vulnerable because slurs are fundamentally less stable signifiers than their neutral counterparts. Signifiers float when unmoored, no longer stable.

Exegesis

Overview

The crux of Neufeld’s analytical strategy is as follows, when she lays out the truth-conditional contribution of slurs:

for a given slur S of a social group G and a person P, S is true of P iff P bears the “essence” of G – whatever that essence is – which is causally responsible for stereotypical negative features associated with G and predicted of P.⁸

Neufeld begins by arguing that “slurs are kind terms encoding an ‘essence’ of a social group, which is taken to explain a number of negative features attributed to the group.”⁹ Now let us consider the slur ‘f****t.’ In Neufeld’s view, this term encodes an apparent essence for all its bearers, those that are said to be ‘f****ts.’ These features obtain in virtue of the bearer’s membership within a socially salient group, which are not there in the slur’s neutral counterpart, the kind GAY MAN. She is

⁸ Neufeld, “An Essentialist Theory,” p. 3.

⁹ Neufeld, “An Essentialist Theory,” p. 2.

insistent that this essence does not actually exist, precisely because there is no actual quality uniting bearers of a slur. The kind F****T does not exist, and bearers of ‘f****t’ do not actually possess some negative quality that is essential to their identity. In other words, though the essence for regular natural kinds such as GAY MAN is hidden, unobservable, or a placeholder notion, there is just no property which x has iff x is a f****t. Any attempt at specifying an essence, its homophobia aside, would lead us back to the regular natural kinds. But slurs are clearly different from those in their evaluative connotation.

Slurs differ from their neutral counterparts in that they are dehumanizing, and this dehumanizing nature is present in all slurs. Dehumanization is therefore at the heart of slur use, even when a slur’s speaker does not use the slur exclusively (one may alternate between ‘f****t’ and ‘gay men’ to describe gay men). Someone may use a slur at a particular moment, while resorting to the neutral counterpart otherwise. Neufeld argues that this is because the two terms are not synonymous, but are related in some way (“counterparts”, “slur-relative”).

Therefore, a slur does not actually refer to its bearers, contrary to what its speaker may believe. For that reason, I will write “essence” in quotes moving forward, to emphasize, much like Neufeld does, that there are no individuals who are essentially less human, contrary to what slurs suggest:

More concretely, there is no such thing as a “gayness essence” which disposes male homosexuals to carry HIV or dress stylishly. There is no such thing as a “blackness

essence” which causes black people to deal drugs or receive welfare. Thus, the semantic contents of slurring words are empty.¹⁰

But though the extension of $F^{****}T$ is empty such that $\{f^{****}t\} = \emptyset$ (hence “failed” natural kind terms), slurs are also not semantic failures. In fact, as Neufeld admits, they are quite “informationally rich.”¹¹ Neufeld proposes that a slur has three core elements. We have so far described the first and most important, which is the “essence”, also the slur’s “causal component.”¹² Slurs differ from evaluatively neutral natural kind terms in that the features they encode (in contrast to neutral features) play a special causal role. A slur’s derogatory “essence”, in contrast to a non-slur’s neutral one, causes the bearer of a slur to be viewed in dehumanized form.

Whilst both ‘ $f^{****}t$ ’ and ‘gay man’ pick out members of some social group G in conventional use – that is, gay men – ‘ $f^{****}t$ ’ is negative not only in connotation. According to Neufeld, it also evokes particularly negative stereotype features of gay men, which is the second core element of slurs: pointing to negative stereotypes that are iconic of the group G . For the homophobe who earnestly believes that there is an essence

¹⁰ Neufeld, “An Essentialist Theory,” p. 3.

¹¹ Ibid. p. 6.

¹² Ibid. p. 3.

undergirding F****T, these stereotypes are proof. Their reasoning posits “essence” as mutually constitutive with what they observe:

- “essence” \Rightarrow “observable surface features”
- e.g. “He’s a f****t, so he has to talk like he’s on helium 24/7.”

as well as:

- “observable surface features” \Rightarrow “essence”
- e.g. “He talks like he’s on helium 24/7, so he’s a f****t.”

Finally, there is the causal law relating “essence” with the stereotypes in the context of utterance, which altogether comprise the slur’s “mini-theory.”¹³ We can see how a slur’s bearer is caught in a double bind, damned if they do not (former), and damned if they do (latter). That is because all slurs presume a static typology of what its bearers must be like, whether or not they actually are, which is why slurs say more about its speakers than bearers. An imagined other cannot talk back.

Neufeld then examines seven different cases of slur usage, and shows how she accounts for each of them, including where

¹³ Ibid. p. 4.

her account differs from those of others. One of the seven is standard derogatory use like the example we have just discussed. She also treats G-extending and G-contracting as two cases, but I will discuss them in tandem for clarity.

G-Extending and G-Contracting

Consider two sets. One is the set containing the bearers of S. The other is the set containing bearers of some evaluatively neutral way to pick out members of G. The difference between G-extending and G-contracting is a matter of which set is a subset/superset of the other. In both, we are presuming that they are not equal, where

- $\{x \mid x \text{ is the bearer of 'f****t'}\} \neq \{x \mid x \text{ is the bearer of 'gay man'}\}$.

In G-extending cases, the slur is the hypernym, such that:

- $\{x \mid x \text{ is the bearer of 'f****t'}\} \supseteq \{x \mid x \text{ is the bearer of 'gay man'}\}$
- e.g. “He’s a f****t, though he’s not gay.”

In G-contracting cases, the slur is the hyponym, such that:

- $\{x \mid x \text{ is the bearer of 'f****t'}\} \subsetneq \{x \mid x \text{ is the bearer of 'gay man'}\}$
- e.g. “He’s not a f****t, though he’s gay.”

This case demonstrates well that slurs are truth-conditionally separate from their neutral counterparts, and as such are essentially different.

Non-Derogatory, Non-Appropriated (NDNA) Use

Neufeld considers a case with the slur ‘chink’ and its neutral counterpart ‘Chinese.’ As she writes, “‘Chinese’ and ‘chink’ are not synonymous according to my account; only the latter term is true of those individuals that share a ‘Chinese essence’ which causes them to exhibit negative stereotypical features.”¹⁴ But NDNA use, as she sees it, is more than just a matter of distancing slurs from their neutral counterparts semantically. We have already established that in the above case.

The difference between NDNA use and G-extending/contracting, as Neufeld argues, is that slurs are used in NDNA cases as part of a broader evaluative statement which does seem to be morally acceptable. Such a use is what cancels (if not completely) their being derogatory. Consider these two statements:

- He has a name, and it’s ‘chink.’ (obviously derogatory)
- He has a name, and it’s not ‘chink.’ (much less derogatory)

¹⁴ Ibid. p. 13.

Null-Extension

Neufeld acknowledges that certain negative stereotypes are sociohistorically grounded. In other words, there is a causal connection between membership in G and possessing negative stereotypes associated with G, mediated by structural injustices. A slur may then pass the causal requirement and refer. One thinks of Richard Wright’s novel *Native Son* as a poignant portrait of how African-American identity, paradigmatic of any racialized identity, is caught in a vicious cycle of being both perpetrator and victim of labels of deviance.¹⁵

But as Neufeld would argue, to identify Wright’s protagonist, Bigger, with a slur would be to suggest that he “possesses some intrinsic ‘[blackness] essence.’”¹⁶ Such a suggestion would be mistaken, as it would be confusing the features – causally rooted in oppression – with the “essence.” The latter, as Neufeld shows, has to be intrinsic, but with Bigger,

¹⁵ Readers charitable to my sociological background may be interested in Howard Becker’s labelling theory (a cornerstone of contemporary social psychology) for a perspective on how slurs cause the slurred upon to see themselves in terms of the slur’s presuppositions, thereby internalizing pejorative affect. This phenomenon is perhaps best thought of as delivering the opposite of the emancipation that comes with appropriated uses. It is perhaps interesting to consider the sociological questions surrounding (in)appropriate use, but to venture farther into the social construction of slurs would be outside the scope of this paper.

¹⁶ Neufeld, “An Essentialist Theory,” 3, pp. 13–14.

as with all African-Americans and racialized people, it is not. It is a social property. There is a plethora of sociological literature on the construction of “blackness” and “whiteness” as an imagined dichotomy *vis-à-vis* colonial hegemony which was first and foremost a moral project. Its goal was to distill into the colonized (in this case “blackness”) a genuine psychological inferiority, as part of what Frantz Fanon termed a Manichaeon world.

She goes on to explain null-extension cases which are trivially true, which need not concern us here.

Derogatory Projection

Derogatory projection refers to the use of slurs in a way such that their existence is presupposed, even if their “essence” is not the focus of the assertion. Such a use is more indirectly derogatory, but Neufeld’s account is still able to deny the presuppositions as false when the slurs themselves fail, if separate from the assertion itself.

Derogatory Variation

Finally, derogatory variation addresses the fact that not all slurs are equal. Rather, slurs vary in degrees of dehumanization, which are accounted for by degrees of essentialization. The most derogatory slurs are the slurs which insult the most personal. At the start of the exegesis, I emphasized specifically the bearer’s membership within a socially salient group. A slur relating to my membership in a group which prefers pumpkin over pecan pie

(‘pumpkinhead’) seems innocuous, because the essentialization of food preference pales in salience compared to that of race, sexual orientation, or other demographic factors.

Interdisciplinary Connections

Neufeld concludes with evidence from cognitive psychology on how humans categorize objects in terms of natural kinds, boundaries between these natural kinds therefore being seen as immutable. More specifically, she mentions how nominalization contributes to a feeling of essentialization as an immutable part of identity. Describing someone as a *x*, where *x* is a noun, contributes much more to *x*-ness being seen as a part of their identity than simply saying that someone is/does *x*-ing.

Objection

Overview

Before delving into slurs in natural language, Neufeld contrasts her account with those of others, notably Elisabeth Camp’s perspectival account.

In contrast, although they can reveal something about my perspective ... slurs are not *about* perspectives on my account. In my view, slurs’ meaning is predicative in the full-fledged, traditional sense. What slurs *say* of you is that you have some group essence that disposes you to display bad features; thus, a slur-predication will be either true or

false of you. It is precisely because of the purely predicative function that slurs come out as empty.¹⁷

But Camp’s account of slurs as a dual act may yet yield some insight. Specifically, Camp’s second line of reasoning that slurs act to endorse “a derogating perspective” promises to capture the nuances of affect which have eluded Neufeld, and supplement her essentialist account well.¹⁸ Though Camp writes that “perspectives are modes of interpretation that structure an overall collection of thoughts in an intuitive, holistic way without necessarily committing to any particular proposition, emotion, or evaluation”, it is clear that slurs are sometimes used as floating signifiers, including when they are used for affective hostility flouting conventions, and perspectives *à la* Camp come closest to accounting for this phenomenon.¹⁹ I will consider two cases.

¹⁷ Neufeld, “An Essentialist Theory,” 7.

¹⁸ Elisabeth Camp, “A Dual Act Analysis of Slurs,” in *Bad Words: Philosophical Perspectives on Slurs*, ed. David Sosa (Oxford UP, 2018), 30, <https://doi.org/10.1093/oso/9780198758655.003.0003>.

¹⁹ Camp, “A Dual Act,” p. 50.

Use as Floating Signifiers: Semantic Bleaching/Widening

This case bears similarities to G-extending and G-contracting, but it differs from the examples Neufeld considers in that the slur does not have a neutral counterpart. Though G exists, the bounds delineating it are undefined. Let us consider a possible world where, due to various mishaps in my upbringing, I frequently utter statements such as the following pejoratively:

- Canada is full of leftist b***hes.

Who are the bearers of S (‘b***hes’) in this case? No inference can be made other than the fact that I have expressed a hostile affect (perhaps contempt, frustration, or a mix of both) towards people in Canada. But I am only to find a salient group G because ‘Canada’ is present explicitly. I am not able to correlate the slur with Canadians of any specific gender, because the gender connotation is not (or perhaps no longer) central to the “essence” of S. I am not even able to correlate S with Canadians that espouse political beliefs left of centre, because I may know nothing about what those beliefs actually are. S (‘b***hes’) is therefore able to extend and contract without reference to its conventional group G (women), becoming unmoored from its conventional counterpart (‘women’).

I propose that Neufeld accounts for this case in the way she does between the blending of slurs and “individual

pejoratives.”²⁰ The slur ‘b***h’ is an example of a slur that has undergone semantic bleaching/widening, becoming less stable. The slur can be more individualistic in that it can be tailored bespoke to the target in question. The target for the speaker ranges on a continuum from anyone (very individualistic, like “jerk”) to women with political beliefs left of centre (less individualistic, and more like a conventional slur), depending on what kind of meaning the speaker wants to express. Neufeld’s approach to slurs along a continuum has the potential to accommodate bearers of slurs along a similar continuum. As I see it, the continuum notion is one of the strongest aspects of her theory, itself an introspection into SLUR as a natural kind. Recall that the central reason slurs fail as natural kind terms is that there is no essence – there is no property which *x* has iff *x* is a b***h. This scenario is a more lucid example of how the slur does not refer.

A consideration of perspective which motivates such an utterance could enable Neufeld’s account to deal with affective hostility more directly. One could argue that slurs are at least to some degree cooperative; they express an affective disposition, and the speaker’s intention is for the bearer to know that they harbour pejorative feelings. The appeal of slurs for the ordinary user is that the term is directed at an imagined category, whatever is contextually desirable at the time. As Neufeld shows

²⁰ Neufeld, “An Essentialist Theory,” 18–19.

with the “queers” example (quoted from Jeshion), it is not hard to picture a child growing up who has never met any LGBTQ+ individuals, who has no idea who they even are, but in the case that they hear family members use the slur disparagingly, they will engage in slurring use as well. Affect can be learnt from prosody alone. Insofar as slurs are inherently political, my hope is to converge ever closer on a politically responsible account of them. Slurs are routinely deployed in political rhetoric as vehicles of demagoguery, and I would be irresponsible to neglect such a reality.

Use as Floating Signifiers: Derogation-cum-Infelicity

This case is not categorizable under any type of use that Neufeld considers. Slurs are much more likely than regular natural kind terms to remain charged in meaning even under inappropriate use by an incompetent speaker. Misattribution of a slur may signal incompetence, but it certainly does not cancel derogative attitudes. Consider a speaker singling out an individual who, in their eyes, displays negative stereotypical features associated with being gay, then uttering:

- What a chink!

The speaker has directed the slur at the bearer in virtue of the bearer’s perceived membership in the socially salient group of gay men, but this utterance is not felicitous because of a disjunction in conventionality linking a given G with its conventional S. Perhaps the speaker is not a competent English

speaker, or perhaps they see Chinese men as displaying the same negative stereotypical features as gay men and thus equivocate the two’s “essences”, but we cannot know for sure. A felicitous utterance signals homophobia, but this utterance does not signal the absence of homophobia. It may actually signal additional racist prejudices harboured by the speaker in addition to homophobia if the speaker is able to grasp the meaning of ‘chink.’

Camp is especially helpful here in asserting that we do not need to know what is going on in the speaker’s psyche, writing that “commitment to the appropriateness of a perspective is not itself a straightforwardly doxastic attitude; in this sense, my view is broadly expressivist, although it abjures any conventional connection to the expression of specific feelings like contempt.”²¹ Currently, Neufeld posits that these two lexical items (‘f****t’ and ‘chink’) differ in terms of the surface features that are part and parcel of their respective causal components. In other words, they differ essentially, in terms of meaning. But this architecture still feels too cumbersome. We want to say that the speaker is at least a homophobe if not a racist, and is using ‘chink’ to express at least homophobia if not racism, a standard derogatory use of slurs in either/both cases. It is not the case that perspective is separate from and inherently extrinsic to meaning. If it is “persons *qua* persons” being impacted, then it makes no

²¹ Camp, “A Dual Act,” p. 50.

sense to oppose some slurs *qua* affects and not others, when the affects are the same.²²

Conclusion

Let us return to the passage which is the site of my objection, the specific at-issue claim italicized for emphasis:

In contrast, although they can reveal something about my perspective ... slurs are not *about* perspectives on my account. In my view, slurs' meaning is predicative in the full-fledged, traditional sense. What slurs *say* of you is that *you have some group essence that disposes you to display bad features* [emphasis added]; thus, a slur-predication will be either true or false of you. It is precisely because of the purely predicative function that slurs come out as empty.

I have shown how the causal connection from the bearer of a slur to the truth value of a slur-predication in the following process may be weakened:

- “essence” ⇒ “observable surface features” ⇒ slur-predication

And I have done so through two cases showing how slurs can “*say of*” something else: 1) semantic bleaching/widening, and 2) derogation-cum-infelicity. In the first case, the bounds of

²² Neufeld, “An Essentialist Theory,” p. 12.

the group G are indeterminate. In the second case, slur-predications can return truth or falsity for bearers not traditionally considered to be in G or in another G , due to infelicitous use.

Therefore, I suggest the following modification to Neufeld’s definition of slurs: for a given slur S of a perceived socially salient group G by the speaker P_1 and a bearer P_2 , S is true of P_2 iff P_2 bears the “essence” of G – whatever that essence is – which is causally responsible for stereotypical negative features associated with G and predicted of P_2 .

The adjustment, though seemingly cosmetic, is important in that it emphasizes the speaker’s role and agency, especially where the slur S becomes detached from its conventional group G , but rather becomes a floating signifier of the speaker’s own (albeit social) molding. Both Neufeld and I have focused on approaching slurs in terms of the speaker’s perspective and intention, in terms of what they see or do not see, even though Neufeld mistakes a similar sentiment from Camp to be incompatible.

I differentiate P_1 and P_2 to affirm that the burden of linguistic labour in slurring language lies squarely with the former, not to suggest a division. The slur-predication has little to do with the bearer P_2 , because the slur is not about any real essence that P_2 may exhibit as part of their membership in real natural kinds. It is our hope that the slurred upon should recover a sense of personhood, made evident when Neufeld writes that “by using slurs ... we commit to a way of carving up the social

world that is dehumanizing.”²³ Hence it is toward a humanizing account that we strive, one that does not fetishize the effect of being slurred upon.

²³ Neufeld, “An Essentialist Theory,” p. 23.

References

- Camp, Elisabeth. “A Dual Act Analysis of Slurs.” In *Bad Words: Philosophical Perspectives on Slurs*, edited by David Sosa. Oxford UP, 2018.
<https://doi.org/10.1093/oso/9780198758655.003.0003>
- Kripke, Saul. *Naming and Necessity*. Basil Blackwell, 1980.
- Neufeld, Eleonore. “An Essentialist Theory of the Meaning of Slurs.” *Philosophers’ Imprint* 19, no. 35 (2019): 1–29.
<http://hdl.handle.net/2027/spo.3521354.0019.035>.
- Putnam, Hilary. “The Meaning of ‘Meaning.’” *Minnesota Studies in the Philosophy of Science* 7 (1975): 131–93.
<https://hdl.handle.net/11299/185225>.

David King

This flash fiction piece is written from the perspective of a modern-day "King David". It explores themes of addiction, human psychology, desire, narcissism, religion, tragedy, and acceptance.

I used to be a cocaine addict.

For nearly three decades, I stumbled and tripped through a murky, yet affable darkness. The giant whose shadow I'd ensconced myself with was both charming and precarious. It took bravery to resist him. To look him in his black, provoking eyes and pull the trigger. When he finally fell, the ground I stood on crumbled and cracked. He almost buried me. But I severed his head. I put it on a pike and left it for the buzzards. Sayonara.

I used to hate women.

I am not proud of this by any means. It's my mother's fault. I think Freud and I would've got on famously. After all, Mother is a boy's introduction to the wondrous world of women. It's difficult then, if at an early age, your mother proves to be deceptive and cruel. It took years to relinquish the hatred I had for women, and for her. I am reluctant to admit this, but I

mistreated every female who entered my life. Now, I fear, I will die alone. Perhaps I deserve it. Who knows? Only God knows.

I used to believe in God.

I prayed fervently. I worshipped regularly. Dancing daily in the streets, arching my back, protruding my chest towards the heavens. He and I were inseparable. I suppose you could say I was a man after his own heart. That was until my son, my golden one, was snatched away by an intruder. An intruder I am still trying to understand, and I'm not convinced I ever will. How is it that God helped my grandmother find a new apartment, yet he ignored the cries of an infant? My infant? Perhaps I'm just being narcissistic. But then again, perhaps not.

I used to be a musician.

A singer. A painter. A poet. It's always intrigued me – creating art. It's magic in its truest sense. Invisible, non-tangible ideas, drawn from the frigid depths of an atom-less psyche. Transformed into something physical. Something with molecules. Something which can be observed, measured, experienced, repeated. Something that matters. Perhaps there's a reason we find Magi in the word Imagination. Words are fascinating. I mean, did Jews invent Jewellery? Now I'm just being ridiculous. As a proud descendent of the tribe of Judah, I apologize...

I'd prefer to begin my tale with the day I met Beth – my ex. Like all exes, she marked the center of my universe. She held more buried treasure in her pinky than any other woman I have ever known. Regardless of our relationship failing, and contrary to what you may have read in the newspapers, we were meant for each other. My corpse will be carted to the mortuary believing that to be so.

The year was 1993. I was living in a penthouse apartment off King St. West. It was mid-July. The air was muggy and dense. I had a breathtaking view of the city. I mean, I could see everything. What I saw that day was unforgettable. A woman, not a day over twenty, swimming in her roof-top pool. Hair: poker-straight, and black. Figure: voluptuous, but firm. She was sculpted by the gods for Christ's sake.

I was close enough to see her bronzed-skin glisten, yet far enough to avoid any blemishes. She was perfect. And, after touring with my band for over a year, I had seen my share of gorgeous women.

Observing her bob, dip, paddle, stroke, back and forth, like some aquatic goddess my mind was made, I had to have her. No matter the cost.

Within forty-eight hours I had her name, her apartment number, her shoe size. I knew where she got her cappuccino; I knew her brand of cigarettes, even her favourite meal.

When you're a rising star, people tend to be drawn to you. I was blowing through copious amounts of cocaine, spending lavishly. The people around at the time, I liken them to corns and bunions—they were always at my feet. They'd do anything I asked, so long as I kept them fueled and fed.

I remember hearing an interview with David Bowie; he was discussing his experience living in Los Angeles—how easy it is to be served in that town. Any thing, any time, at your service, sir. That's how it was for me as well. When I set my sights on something, it wasn't long before I had it. However, with regards to Beth, there was a slight problem. She was married.

So, I had her husband murdered. Seriously. For one gram of cocaine and a bottle of blended Scotch, his throat was slit, and he was left in a heap. His entire history leaking out of his body, seeping into the tarmac. Every laugh. Every quarrel. Birthdays. Bar-Mitzvahs. All of it.

Eliminated.

My execution is now scheduled for April 5th, and I feel that I am ready. A man such as I will inevitably find it difficult to fathom nonexistence, but I suppose this is due to narcissism. For out of the billions of humans on this cruel planet, more than ninety percent are blissfully unaware that I ever was. Why then should it be so inconceivable, so harrowing for *me* to accept?

I now realize the ego is an illusion. There is no tiny man between my ears or behind my eyes. Perhaps when we awaken to this fact, when we willingly sacrifice ourselves, it is *then* we can truly transcend, rise and live forever. When one accepts, they are not a separate entity, that underneath it all, boundaries are nothing but a figment of the imagination, that humankind is a product of the cosmos, not a foreigner rotting apart from it, well then, it is quite impossible to ever really disappear, isn't it? And, therefore, Robert Plant had it when in a shrill falsetto he wailed: 'All is one and one is all'. It is in realizing *this* that we can finally ascend on our stairway to heaven.

The Monster

A poem on the beginnings and ends of addiction.

Sugar, spice, everything nice
There's a price to every vice
Be your family, health, or friends
One or more will meet their ends

The beginning is pure bliss
It's a feeling you will miss
Leave your world and mind behind
Find joy of a different kind

Perhaps the monster was always there
Watching you with great care
Waiting until your first toke
When you trade your life for smoke

The monster is hungry, and it's patient
It will return when you're complacent
It feeds on pain and loss and fear
When you're alone, it draws near

That first toke unlocked the door
The monster is hungry; it wants more
The second toke beacons it inside
Once it's there, you cannot hide

Your closest friend, your deepest love
No one will ever come above
Not your children, life, or spouse
You'll even give that guy your house

I wish I could return to the beginning
When life without the monster had some meaning
When my deepest love and closest friend
Was a book with a happy end

I can hear the monster roar
Scratch outside the bedroom door
It will be there till I'm dead
But hopefully, not in my bed

The Philosophy of Environmental Design

My creative piece, "The Philosophy of Environmental Design", advocates for the reader to reconsider what the places they encounter or live life through say about themselves through visual design. The creative prose asks the reader if a living space should not just be a place to live, but rather a place to self-express. This short piece of writing emphasizes the importance of acknowledging past and future ambitions through visual design to help enhance reflection in the present moment.

I do a lot of travelling, a lot of visiting, from the start of my day until the end of it. When I am allowed to visit a friend or get invited to a coffee shop, I find my mind questioning the environment design of the places I visit. I question why the environment I live in is designed in such a way. And what the environment says about itself, its owner, or its own philosophy through its built design.

When I visit a friend, I look around at the small details of their house. Not because I am nosey, but rather because I question how their interior design reflects them as a person. Do they have books because they are academics? Or are they just avid readers? Is the small car model on their table just a gift? Or is it a reflection of a passion or hobby? I believe that if we're to

consider the environment we inhabit a conversation piece, not just a place to wind down, the internal design philosophy of the home then changes. It becomes a personal canvas for self-expression. Even staying on trend, or designing the living space in order to accommodate the trend speaks for itself. It then switches the self-expressive home into a member towards what the design trend represents, rather than a self-expressive space.

The same goes for corporate businesses. I find that with corporate buildings, the design is based on modern trends rather than a specific philosophy. It makes those short visits to family-owned businesses all the more exciting for someone like me. The corporation is built on adaptation, yet the home is built upon a journey.

I find myself distancing my living space from basic trendiness or order. While I still believe a space should be clean, of course, I aim to include myself within the space between the symmetry of my furniture. I aim to be the designer of my living space in a way that is reflective of my journey, my memories, and the people I cherish throughout life. I argue that personal expression through our materialistic possessions and living spaces enhance and re-contextualize who we are, what we want to be, and how we want others to perceive us. By allowing ourselves to be expressed this way, we provide ourselves with a journey and a lookback from the start till the present moment.

Even if you love orderly design, or prefer minimalism, somewhere in between, don't forget to add a bit of yourself. Your past and future self will thank you for it.

Rebirth of the Forest

Destruction leads to new beginnings. This is a photo of an area near Bootski Lake, British Columbia, that was destroyed by a wildfire in 2006. The charred and leafless trees, through which you can see the mountains, juxtapose the lush green bushes and Castilleja growing on the forest floor. This shows that after an ending, a rebirth can occur. Destruction creates space for renewal, which reveals that a loss means a change, not an end. The regrowth does not erase the damage caused by the wildfire, but it can create hope for a new life. The forest is still a forest and, even after a loss, it remains.

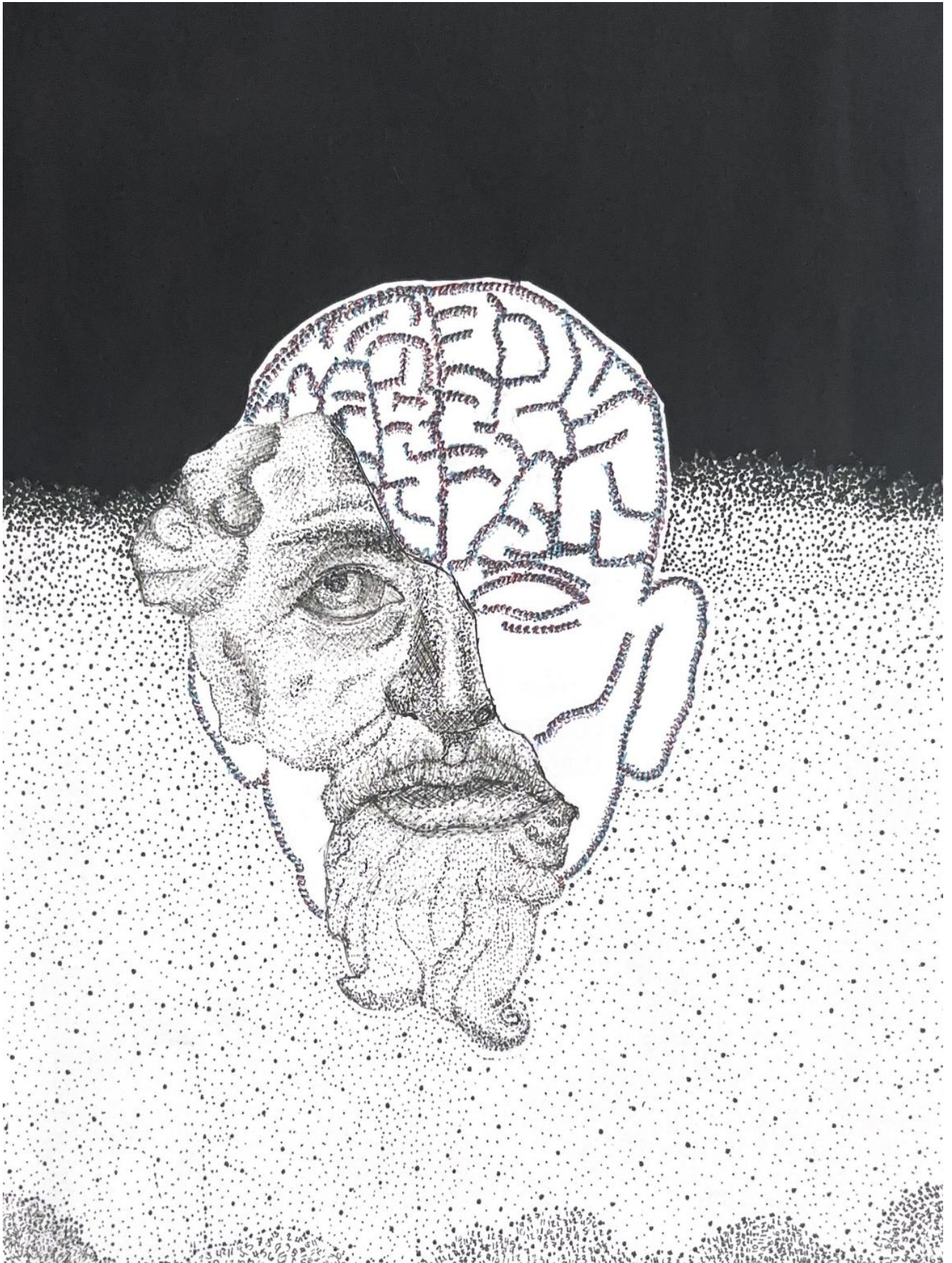


The Questions Inside of Us

This piece is inspired by Marcus Aurelius, who was a stoic philosopher and former Roman emperor. His face is broken open to reveal a brain and human anatomy made entirely of question marks.

Even after everything we have learned and everything we think we know, we find ourselves unsure of who we are and are left with many unanswered questions. The stippling effect represents the passage of time. It shows how the body slowly wears down while the mind searches for meaning. The solid black background fades into stippling and becomes clouds made up of the numbers one through twelve, referencing a clock and the movement of time. The red and blue question marks add a glitch effect and a small pop of colour, suggesting moments of distortion within our thoughts.

This piece reflects on life from our first breath to our very last. It looks at how we move through the world carrying questions we may never have answers to; this shapes who we are, the choices we make, and the meaning we try to build before everything eventually fades away.

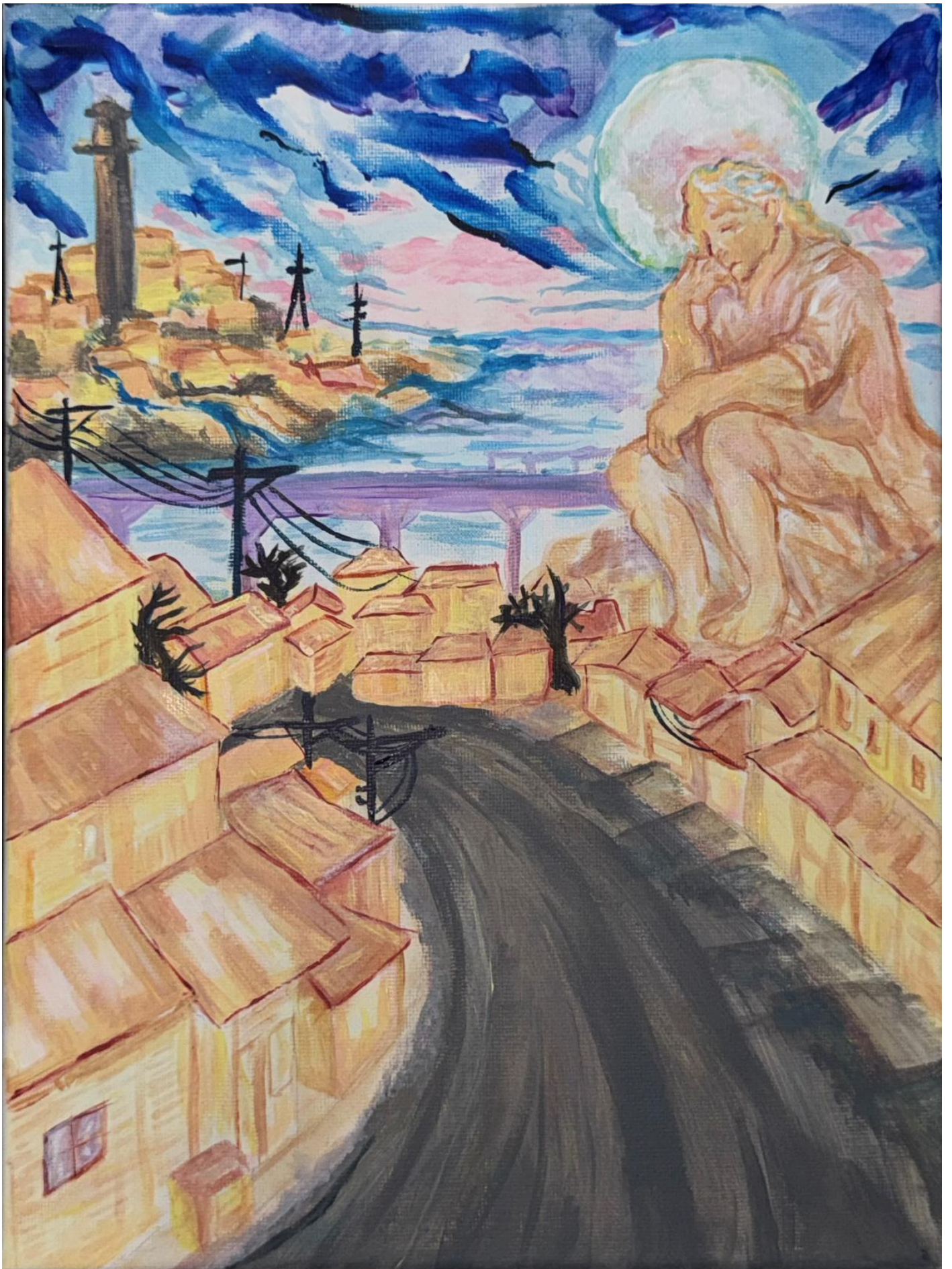


The Witness

A great many disciplines have emerged, originally, from philosophy. It is a subject that does not always take the forefront, but incubates others and watches as they rise into public interest or fall into obscurity.

The city represents the myriad systems philosophy has given rise to. The clustered buildings, wires, and structures feel accumulated, while the road encourages movement forward in time.

The Thinker by Rodin stands as a primary representation of philosophy in popular culture. This work treats The Thinker as both a beginning and an end. The figure is monumental yet eroded, suggesting a presence that has been repeatedly drawn from, fractured, and repurposed. Philosophy persists not as resolution, but as witness, remaining after certainty has faded and before a new inquiry begins.



Untitled

This is an abstract portrait painting I made as a short passion project last summer. It represents the non-corporal nature of philosophical thought.



