The Brown Journal of Philosophy

A PRIORI

Volume 7

Brown University

Department of Philosophy

2024

Brown University Undergraduate Journal of Philosophy

A Priori

Volume VII

Acknowledgments

A Priori would like to thank the Brown University Philosophy Department for its continued support and funding. Additionally, we would like to thank the numerous students and universities that submitted papers from across the world and shared their ideas with us. We are grateful, too, to the professors who have guided all of us in our love of philosophy and search for knowledge.

MASTHEAD

Editors-in-Chief:	Sophia Decherney Jaehyun Hong Julie Qian
Cover Design:	Becky Montes
Editorial Board:	Lucy Cooper-Silvis Daniel Kang San Kwon David Tapper Katherine Warren
Editors:	Alex Anaya William Hassett Peter Haynes Koda Li Charles Luke Zoya Unni

Contents

An Examination of the Contemplative Life and Social Relationships in Nicomachean Ethics

Simeng Wang

Aristotle has contributed tremendously to the realm of moral ethics by theorizing the happy and fulfilling life. In his representative work, Nicomachean Ethics, he explains the role virtue plays in consolidating that kind of life. While scholars have conducted countless kinds of literature in understanding how significant having an excellent character is towards achieving well-being, there seems to be a lack of interest in interpreting Book X of NE, which specifically talks about a contemplative life built on self-sufficiency and meaningful social relationships. Although the two concepts appear contradictory on the surface, in this essay, I argue that they are neatly coherent and compatible with each other. I aim to clarify this subtle relationship to further emphasize how practical and well-thought-out Aristotle's philosophy is, and consequently, why it deserves to be read and examined by us to this very day.

In *Nicomachean Ethics* Book X, Aristotle writes that the most satisfying and fulfilling life is a life of study, one based on continuous and reflective contemplation. He does, however, acknowledge that for humans to thrive in communities as social and political beings, they must maintain relationships with other people in order to gain insights from the external world. Are these two points contradictory, since social interactions might interfere with one's focus on introspection? Or are they compatible with one another if we recognize the nuanced connection between them? In this essay, I want to discuss how social relations, especially friendships that are complete, noble, and mutually beneficial, are essential for a self-sufficient life of study for ordinary people.

The initial step to solving the above-mentioned puzzle is to have a clearer picture of what Aristotle means when he refers to a contemplative life. I offer two perspectives for understanding the contemplative life. The first is to recognize the pleasant and self-sufficient nature of this kind of life. Book X describes pleasure as being highly context-related. Pleasure is not freestanding but is rather an epiphenomenon that comes along with distinct activities. Besides that, Aristotle also points out that certain pleasures are derived from the replenishment of preexisting emptiness, such as the joviality we feel when we consume food to drive out the pain of hunger.[1] In contrast, a contemplative life is not a stuffing process, "since no emptiness of anything has come to be, there is nothing whose refilling might come to be."[2] It contains a pure and firm pleasure that engages with a person's higher-order faculties. Furthermore, a contemplative life follows the formal criteria that Aristotle has been developing throughout the whole book. A contemplative life's value is not "derivative upon certain external ends."[3] Engaging oneself in this life is pleasurable and whole simultaneously. One continuously finds joy since contemplation is serious yet enjoyable. At the same time, this life which is "superior in

[1] Aristotle, *Nicomachean Ethics*, trans. Terence Irwin (Indianapolis: Hackett, 2019), 185.
[2] Aristotle, *Nicomachean Ethics,* 185.
[3] Nancy Sherman, *The Fabric of Character: Aristotle's Theory of Virtue* (Oxford: Clarendon Press, 1989), 98.

excellence” has all its value fully reserved within itself rather than on anything further or external to it.[4]

After understanding the structural features of the contemplative life, one might wonder about its substantial content. What exactly does a contemplative life refer to? The life devoted to *theoria* ought to be lived by “someone who has a full understanding of the basic causal principles that govern the operation of the universe, and who has the resources needed for living a life devoted to the exercise of that understanding.”[5] That is to say, the contemplative life Aristotle refers to ought to be carried out toward certain overarching schemas that are built upon—yet surpass—a simple discernment of mundane reality. Hence, contemplation as a method to get closer access to the unchanging nature of the world is not tantamount to deliberation. Contemplation enables one to be conscious of one’s thinking patterns from a more holistic standpoint instead of focusing solely on the subject of thinking. Aiming toward the eternal truths of the universe and the sanctity of life, contemplation is rumination toward the “fine and divine.”[6]

Nevertheless, can normal human beings adapt to this way of living that transcends mediocrity? Is this life a bliss that ordinary people could even have a chance to experience? Aristotle addresses this inquiry later in Chapter 7 of Book X. He first admits that the

[4] Aristotle, *Nicomachean Ethics,* 194.

[5] Richard Kraut, “Aristotle’s Ethics,” *The Stanford Encyclopedia of Philosophy*, The Metaphysics Research Lab at Stanford University, updated Fall 2022, https://plato.stanford.edu/archives/fall2022/entries/aristotle-ethics/.

[6] Aristotle, *Nicomachean Ethics,* 193.

contemplative life he mentions “would be superior to the human level.”[7] Insofar as one remains a human being, there will be innate obstacles stopping one from fully reaching that kind of life since one does not possess all the necessary divine elements. However, Aristotle also states that “rather, as far as we can, we ought to be pro-immortal, and go to all lengths to live in accord with / our supreme element; for however much this element may lack in bulk, by much more it surpassed everything in power and value.”[8] In short, there are hindrances in human nature that make the perfect life of contemplation unapproachable. But we should treat that life with respect and yearning since it deserves to be treated as such. In other words, the most ideal paradigm of the contemplative life seems unreachable, yet this fact doesn’t render Aristotle’s philosophical discussion void. Instead of debating whether it is ever possible for ordinary human beings to gain access to an unworldly life orientated toward understanding, we should rather discuss the way to reach a contemplative life “not as a god would, but as a human would, with the boundaries defined by our social and moral lives.”[9]

Following that, the remaining task in the rest of the essay is to figure out how humans can live a life of contemplation within their own stretch of capability “*as far as is possible.*”[10] It is not an either yes or no question about whether humans should pursue a standpoint that is “unconditioned by historical circumstances.”[11] What should really be taken into consideration is

[7] Aristotle, *Nicomachean Ethics,* 195.
[8] Aristotle, *Nicomachean Ethics,* 195.
[9] Sherman, *The Fabric of Character,* 101.
[10] Aristotle, *Nicomachean Ethics,* 194.
[11] Sherman, *The Fabric of Character,* 101.

the extent to which humans could navigate toward the life that "allows for a most perfect form of happiness and self-perfection."[12] If it is intrinsically implausible for human beings to reach the life of the philosophy king upon the hill, then how should they shape a way of living that has lofty connotations in a more transcendental dimension while remaining suitable for themselves? Aristotle's perspective on the external resources contributing toward *Eudaimonia* will be particularly helpful in response to these theoretical conundrums.

To clarify, Aristotle never denies the necessity of external resources for flourishing as individuals. A fully prosperous life requires a set of abundant resources. Besides certain basic needs such as shelter, food, and things alike, social connections are the proper mediums for a person's cultivated and habituated virtues to be articulated and implemented into real life. Aristotle states that even a wise person is still in need of other people "as partners and recipients of his just actions."[13] A person ought not to live in total solitude, as external apparatuses based on socially constructed bonding and interactions can participate in one's life in meaningful and influential ways that one cannot generate and sustain on one's own.

Nevertheless, the life of study proves that it is likely for a virtuous person to gain happiness with a relatively small extent of dependency on externalities since a contemplative life requires fewer extrinsic goods to a lesser degree. That does not mean a person devoted to contemplation does not need those resources to flourish. It rather refers to the idea that external

[12] Sherman, *The Fabric of Character,* 101.
[13] Aristotle, *Nicomachean Ethics,* 194.

goods are necessary, but only necessary to a certain extent. One's life will be shaky if one places excessive weight on things out of his scope of control. "Needing necessary resources" and "believing that one can only be well off with those resources" are intrinsically two very different mindsets, and the uniqueness and praiseworthiness of a contemplative life is that it admits the former but rejects the latter. It accepts the irreplaceable status of external resources in one's life, but it reinforces that contemplation needs to be generated from self-sufficiency that is solid and stable on its own behalf. A successful way of living contemplatively is to be willfully ignorant toward the roles external resources play in a person's life: their importance is being recognized and appreciated without generating inordinate dependence on them. One's state of happiness should not be unduly determined by the extent of plentifulness of external resources one can attain in a lifetime, that is, "we must not think that to be happy we will need many large goods."[14]

Undoubtedly, contemplation is often abstract and theological. But one cannot achieve that phase if one has not immersed oneself in cultivating the corresponding virtues in particulars. What role does friendship play within this stretch between the down-to-earth and the divine? In Book VIII, Aristotle differentiates among three kinds of friendship: friendship based on utility and expediency, friendship based on easily dissolvable passion and pleasure, and complete friendship based on both parties' similarities in virtues. Later on in Book IX, Aristotle further elaborates on the desirability of the complete friendship compared to the other two. For a base

[14] Aristotle, *Nicomachean Ethics,* 197.

person, it might be extremely difficult for him to have complete and noble friendships since he is at odds with himself. He is constantly distracted by things that are not constructive toward his personal well-being. He is filled with internal conflicts, therefore he cannot have genuine friendships since he cannot even appear loveable to himself, as "the bad man does not seem to be amicably disposed even to himself, because there is nothing in him to love."[15] In comparison, a good person must be a self-lover since he has fine and harmonious desires guided by reasons and beneficial to both him and others. Furthermore, he can also derive happiness and satisfaction from observing pleasant behaviors of his intimate ones and categorize those behaviors as conduct he is capable of performing as well. In this way, a decent person is related to his friend as he is to himself, since a friend is a lively reflection of the virtuous qualities one already possesses and the potential ones one can foster accordingly. A complete friendship is thereby characterized by similarities in virtue from both parties and a reciprocal goodwill for the other person to flourish, thus consolidating this kind of human relationship to be truly equal, noble, and pleasurable. Aristotle tactfully summarizes this contrast that "the good man should be a lover of self (for he will both himself profit by doing noble acts, and will benefit his fellows), but the wicked man should not…following as he does evil passions."[16] A complete friendship thus becomes the mirror one observes one's own being while offering opportunities for one to build bonds that extend beyond one's inner self and stretch into the outside world.

[15] Aristotle, *The Nicomachean Ethics*, trans. David Ross (New York: Oxford University Press, 2009), 170.
[16] Aristotle, *The Nicomachean Ethics,* trans. David Ross, 175.

I will go on to discuss in further detail the role of friendship, as mentioned in the text, "but the wise person is able, and more able the wiser he is, to study even by himself. Though, presumably, he does / it better with colleagues, even so he is more self-sufficient than any other [virtuous person]."[17] The simple argument that a highly wise person will be able to study better with others' company in life despite the fact that he is truly self-sufficient makes me ponder. Even though this person can generate ample vigor and momentum to keep supporting his intellectual pursuits, Aristotle still senses others' value and efficacy in bettering this life of study. One possible way of understanding is that locking oneself in total solitude and isolation carries the risk of creating an information cocoon, where one doesn't gain a lively and sincere exchange of ideas, evidence, and propositions but rather resides in a bubble of self-justification. Consequently, one might not be able to combat the impact of personal prejudice and preference, which might negatively impact one's grasp of the fundamental truth. Others' company could either function as resonance or logical challenges toward one's held beliefs, thus extending the breadth and depth of one's thinking. Therefore, self-sufficiency and external relationships do not need to be mutually exclusive concepts. A person can preserve his self-sufficiency through contemplative activities that do not posit toward any other ends, while at the same time receiving energy and reflections from other communal beings. Social relationships do not exist to deteriorate the pureness and integrity of one's self-sufficiency if they are as complete as Aristotle says. Quite the contrary, positive social links affirm the elements that constitute a virtuous

[17] Aristotle, *Nicomachean Ethics*, trans. Terence Irwin, 194.

person. They are healthy sources of deepening one’s self-sufficiency, therefore stabilizing one's self-awareness. All these steps are primarily important in constructing full-fledged practical wisdom, the concordance between theoretical reasoning (i.e., contemplation), and practical reasoning. Thus, a contemplative and self-sufficient life is never detached from relationships among different social agents, just as we as normal human beings are never distant from the outside atmosphere that encompasses us.

Works Cited

Aristotle. *The Nicomachean Ethics*. Translated by David Ross. Oxford: Oxford University Press, 2009.

Aristotle. *Nicomachean Ethics*. Translated by Terence Irwin. Indianapolis: Hackett, 2019.

Kraut, Richard.“Aristotle’s Ethics.” *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab at Stanford University. Updated Fall 2022. https://plato.stanford.edu/archives/fall2022/entries/aristotle-ethics/.

Sherman, Nancy. *The Fabric of Character: Aristotle’s Theory of Virtue*. Oxford: Clarendon Press, 1989.

Buddhism, Non-Human Animals, and Selfhood

Tyler Jungbauer

I argue that there is no necessary conceptual reason against attributing the same kind of selfhood to non-human animals as is ascribed to human beings, because we can meaningfully ascribe selfhood to non-human animals if we draw upon the Buddhist deflationary account of selfhood. I begin by outlining our intuitive concept of selfhood as is ascribed to human beings. Then I provide a Buddhist argument against ascribing this intuitive concept to human beings to suggest that we should consider alternative accounts of selfhood. To this end, I briefly describe the Buddhist deflationary account of selfhood — on which being a 'self' consists in being a 'person,' which is a conventional functional, folk psychological concept, unlike our intuitive concept of self. Using the Buddhist view, I give a tentative operational definition of selfhood. Finally, I provide empirical evidence that suggests that members of some non-human species may satisfy this definition and thus be selves in the same sense in which human beings are.

1. Introduction

Prior to philosophical analysis, both philosophers and non-philosophers tend to think of human beings pre-theoretically or intuitively as being 'selves,' or subjects who are causally unconstrained by the world. We take ourselves to be distinct from both our mind and body, to be rather what *owns* or *has* a mind and body — the kind of thing that persists across a whole lifetime despite significant changes in both mind and body.[18] In contrast, both pre-theoretically and post-theoretically, we seem to deny that other animals are 'selves' in *any* sense. However, in

[18] Garfield, *Losing ourselves: Learning to live without a self*, 2-4, 30; Siderits, *Buddhism as philosophy: An introduction*, 32-3.

this essay I argue that we can meaningfully ascribe selfhood to non-human animals.[19] If we draw upon the Buddhist deflationary account of selfhood, then we have *prima facie* reasons to attribute selfhood to some animals.

Altogether, I argue that there is no necessary *conceptual* reason against attributing the same kind of selfhood (properly understood)[20] to animals as we ascribe to humans. For the sake of this paper, I take it that we have a *conceptual* reason against applying our concept of selfhood (properly understood) to animals just in case our concept of selfhood is such that it would be a *category error* to ascribe this concept to animals, insofar as animals are not the right *kind* of entity to which this concept properly applies. Instead, I argue that, if any reasons against attributing full-fledged selfhood to animals exist, then these reasons must be empirical, not conceptual.[21]

I begin by outlining our intuitive concept of selfhood. Then I provide a Buddhist argument against attributing this concept to human beings to suggest that we should consider alternative concepts of selfhood. As an alternative, I describe the Buddhist deflationary account and provide a tentative operational definition of selfhood based on this account. In conclusion, I

[19] Hereafter, I use ‘animals’ to refer to ‘non-human animals.’

[20] By “selfhood (properly understood),” I mean a non-intuitive concept of selfhood that has been developed through philosophical analysis, rather than our intuitive, pre-theoretic concept of selfhood.

[21] Let me clarify further what my thesis is. Given some concept of selfhood C, we have a conceptual reason against attributing C to animals just in case subsuming animals under the extension of C would entail a category error. In this paper, I argue against the claim that there is *no* concept of selfhood C* such that attributing C* to animals, or subsuming animals under the extension of C*, would *not* entail a category error. Consequently, I am arguing that there is *some* concept of selfhood C*—namely, the Buddhist concept of personhood considered as a non-intuitive concept of selfhood—such that attributing C* to animals, or subsuming animals under the extension of C*, would not entail a category error. This is a more precise formulation of my thesis.

provide some empirical evidence that suggests that members of some species may satisfy this operational definition and thus be selves in the same sense in which we are.[22]

2. Our Intuitive Concept of Selfhood

We can explicate selfhood in various ways insofar as different concepts of selfhood exist. A concept of selfhood answers the question "What am I?".[23,24] Both philosophers and non-philosophers seem to share an intuitive, pre-theoretic concept that provides a response to this question. I propose that our intuitive concept of selfhood consists in the following: We take the pronoun 'I' to refer to the self as the subject of experience and agent of action, and we also take the self to be the ontological ground of a person's identity over time.[25] As some numerically identical thing enduring throughout a person's lifetime, the self is unitary, possessing both synchronic and diachronic identity.[26] Consequently, the self is essential to a person, where 'person' denotes some psychophysical complex of mind and body enduring over time in virtue of

[22] One caveat. Our *intuitive* or *pre-theoretic* notion of selfhood, not a philosophically sophisticated notion of such, is the primary object of critique in this paper. (More generally, though, I am interested in arguing that it is false to think that there is no concept of selfhood such that we can attribute this concept to other animals. See footnote 21.) Indeed, the Buddhist deflationary view is a philosophically sophisticated view of selfhood developed in response to the problems arising for our intuitive view. Nevertheless, I do not argue in favor of the Buddhist account. Rather, I draw upon this account only to show that no necessary conceptual reasons prevent us from ascribing the same kind of selfhood to animals as we ascribe to ourselves. (By drawing upon the Buddhist deflationary view, I show *prima facie* that there is *some* concept of selfhood such that we can attribute this concept to other animals in addition to human beings.)

[23] Ganeri, *The self: Naturalism, consciousness, and the first-person stance*, 35.

[24] For arguments in favor of the Buddhist deflationary account of selfhood, see Garfield (2022). For a taxonomy and general overview of various philosophically sophisticated accounts of selfhood, including the Buddhist account(s), see Ganeri (2012).

[25] Siderits, *Buddhism as philosophy: An introduction*, 33; Siderits, *Personal identity and Buddhist philosophy: Empty persons*, 29.

[26] Siderits, *Buddhism as philosophy: An introduction*, 35.

the self.[27]

Garfield (2022) identifies four essential properties of our intuitive concept of selfhood: A self has *priority*, *unity*, *subject-object duality*, and *agency*.[28] A self has *priority* because it has a kind of existence more fundamental than, or ‘prior to,’ that of mind and body: A self is the kind of thing that *owns* or *has* a mind and body, and it is the kind of thing *that experiences* and which would exist even without experience. The self also exemplifies *unity* because it is a unitary thing, not a multiplicity: ‘I’ refers to a metaphysical simple, not a complex of entities or processes. Additionally, the self constitutes the subjective, internal pole of a *subject-object duality*, whereas objects in the world constitute the objective, external pole of this duality. In this way, the self is an internal entity, opposed to all external entities existing in the world (including *other* selves). Finally, the self is thought to be the *agent* who is causally and morally responsible for action. As such, the self is causally unconstrained by the world and, thus, radically free and autonomous.[29]

The question arises whether we can attribute our intuitive concept of selfhood to other animals. The answer seems to be negative. Dornbach (2023) grants that “higher animals” have a “rudimentary selfhood,” but nevertheless maintains that complete or consummate selfhood is unique to humans.[30] Bekoff (2003) makes a similar claim, contending that other species may possess “body-ness” or “mine-ness” (a proprioceptive awareness of one’s body or body parts in

[27] Ibid, 32.

[28] Garfield, *Losing ourselves: Learning to live without a self*, 28.

[29] Ibid, 33.

[30] Dornbach, “Animal selfhood and affectivity in Helmuth Plessner’s philosophical biology,” 225.

space), but not full-fledged “I-ness.” Saidel (2018) also concludes that while many animal species have rich mental lives, they lack any concept of self, which only humans have. The common basis for these claims seems to consist in the proposition that there is some categorical difference, or a difference in kind, between humans and animals that precludes subsuming members of other species under the extension of the concept of ‘self.’

Nevertheless, even if we cannot attribute *consummate* selfhood to animals, perhaps we can still attribute some rudimentary approximation of such to animals, such that both humans and animals nevertheless count as being ‘selves.’ Whether this is so depends on how we cut the pie. If humans are ‘full-fledged selves’ and animals are ‘rudimentary/approximate selves,’ then since both sub-categories fall under the general kind ‘self,’ both humans and animals are ‘selves.’ However, even if we cut the conceptual pie this way, we can nevertheless just as easily say that there is a difference in (sub-)kind between humans and animals, since animals are not ‘full-fledged selves,’ but only ‘approximate selves.’ Ultimately, it seems that animals are simply not the right kind of entity to which the *full-fledged* conception of selfhood applies because they are not unitary subjects of experience and uncaused, autonomous agents possessing mind and body.

It is *this* point that I suggest is misguided.[31] I suggest that this presumed difference in kind, or even sub-kind, is mistaken, and I argue that, instead, conceptual reasons like those

[31] More generally, I suggest that the proposition that there is *no* concept of selfhood that can be ascribed to both human and non-human animals is false and mistaken. See footnotes 21 and 22.

suggested above need not preclude animals from being selves in the *same sense* in which human beings are selves. To this end, I first argue that our intuitive concept of selfhood is mistaken in important ways, and that the concept of self that best answers "What am I?" is not our intuitive concept of selfhood. Then I draw upon the Buddhist deflationary account to show that conceptual reasons need not bar members of other species from being full-fledged selves, insofar as we have available an alternative concept of selfhood that *prima facie* provides a better response to the question "What am I?" than does our intuitive concept.

3. A Buddhist Critique of Our Intuitive Concept

In this section, I adduce a Buddhist argument against the existence of the self as it is intuitively understood, and in the next section, I outline the Buddhist deflationary account that is meant to supplant this intuitive view of the self. While offering the argument below, I do not suggest that this argument is conclusive. I only suggest that it is both plausible and counts as a *prima facie* reason against our intuitive view. Given a plausible deflationary account of selfhood and the problems to be identified for our intuitive view, the proponent of this view has the burden of proof to show why we should favor her view over the deflationary one.

Buddhists grant that we have an intuitive self-concept but deny that this concept actually captures what it means to be a human being.[32] Consequently, Buddhists reject that the 'self' as intuitively understood exists. The Buddhist philosopher Nāgārjuna (c. 150 CE?) provides the

[32] Ganeri, *The self: Naturalism, consciousness, and the first-person stance*, 31; Rahula, *What the Buddha taught*, 20–28; Siderits, *Buddhism as philosophy: An introduction*, 35–37.

following reasoning for the conclusion that the self, as it is intuitively understood, does not exist.[33]

Nāgārjuna suggests that if we countenance selves, as intuitively understood, in our ontology, then we must specify how the self so understood is related to the psychophysical complex constituting a person. (A psychophysical complex is (roughly) just a bundle of mental states and physical states at a time, including thoughts, sensations, and bodily processes.[34]) The self can be related to the person in two ways. The first view is that the self is identical with, or reducible to, the psychophysical processes constituting a person. This view is *Reductionism*. The second view is that the self is irreducible to the psychophysical processes constitutive of a person, but is nevertheless related to these processes in some specifiable way. This view is *Non-reductionism*.[35] Nāgārjuna argues against both views.

First, consider Reductionism. Nāgārjuna suggests that the nature of the self as it is intuitively understood is inconsistent with the nature of the psychophysical processes comprising its (putative) reduction base. According to Reductionism, the self is identical with, and reducible to, (some proper part of) the psychophysical complex constituting a person. If the self is identical

[33] MMK, XVIII.1.

[34] See Rahula, *What the Buddha taught*, 51–66, for a more careful examination of what Buddhists take the nature of the constituents of a psychophysical complex to be.

[35] Reductionists and Non-reductionists disagree about which kinds of entities constitute our ontology. Non-reductionists take 'selves' to be part of our ontology because, they claim, selves cannot be reduced to psychophysical processes. Reductionists deny this point on the grounds that we can reduce 'selves' to more basic psychophysical processes, which instead comprise our ontology. Nevertheless, Reductionists do not deny that selves exist *simpliciter*: 'selves' simply consist in the existence of these more basic psychophysical processes (Siderits, 2015, 9–10).

with, and reducible to, the psychophysical complex, then the identity conditions for the self must be the same as the identity conditions for the psychophysical complex.[36] However, while a psychophysical complex has synchronic identity, it lacks diachronic identity. This is because psychophysical complexes are impermanent bundles of mental and physical states that change over time.[37] In contrast, the *self*, by its nature, possesses diachronic identity.[38,39] Therefore, the self and any given psychophysical complex differ in their identity conditions. If identity is necessary for reduction, then it follows that the self cannot be reduced to any psychophysical complex. Altogether, Nāgārjuna argues, Reductionism fails. (The same kind of argument can be used to show that the self cannot be reduced to any *proper part* of some psychophysical complex.)

Reductionism takes the person to be nothing more than some psychophysical complex.[40] Non-reductionism denies this exhaustiveness claim to hold that the person consists of *both* some psychophysical complex and some irreducibly distinct constituent. This additional constituent is the self.[41] If the self is some *sui generis* entity, then the self is not identical with any psychophysical process (or set of processes), in which case the self must instantiate some kind of

[36] By 'identity conditions,' I mean the conditions that define some entity's identity. One version of Leibniz's law states that $x = y$ iff (a) every predicate P of x is a predicate of y and (b) every predicate Q of y is a predicate of x. Conditions (a) and (b) specify the identity conditions for x and y. Thus, if the self is identical with some psychophysical complex, then every property of the self must be a property of this complex, and vice versa.
[37] Siderits, *Buddhism as philosophy: An introduction*, 37–46.
[38] Siderits, *Personal identity and Buddhist philosophy: Empty persons*, 30.
[39] Indeed, this is why we appeal to the self to explain the personal (diachronic) identity of psychophysical complexes, or persons, over time (see Siderits, 2007, 32–33).
[40] Siderits, *Buddhism as philosophy: An introduction*, 50.
[41] Ibid, 32.

property not instantiated by any psychophysical complex. A psychophysical complex instantiates physical and psychological kinds of properties. Hence, the self must instantiate some kind of property that is non-physical and non-psychological. Call this kind of property a 'NN property' (for 'non-physical and non-psychological'), in contrast to a 'PP property' (for 'physical and psychological').

The problem with Non-reductionism is that it is unclear what the relevant NN properties would consist of. Presumably, the only kinds of properties relevant to specifying the relationship between the self and some psychophysical complex are those discoverable in experience. (Otherwise, it is unclear how we would know the properties in question.[42]) However, the kinds of properties discoverable in experience are PP properties, not NN properties. Furthermore, even if we grant that NN properties exist, we still must explain how the self, which instantiates NN properties, can causally interact with the physical and psychological part(s) of reality, which instantiates only PP properties. Such an explanation must answer two questions: (a) What kind of causal relations hold between NN properties and PP properties in virtue of which selves can causally interact with the physical and psychological part(s) of reality? (b) If we can explain all causal phenomena involving PP properties without positing NN properties, then why should we

[42] Perhaps we know NN properties by *a priori* intuition. While plausible, I am unconvinced by this suggestion. If we know NN properties by *a priori* intuition, then that NN properties exist is a necessary truth. However, that NN properties exist does not seem to be necessarily true at all. What is necessarily true is what is true at all possible worlds. Certainly, though, we can imagine possible worlds at which it is false that NN properties exist; indeed, that the self exists seems to be a contingent, non-necessary matter. More precisely, we can imagine possible worlds at which it is false that the NN properties *relevant to the self* exist, even if we want to grant that for all possible worlds, it is true that there exist *some* NN properties—just not those NN properties that are relevant to the existence of the self.

posit any causal relations that would answer (a)? If we can explain all causal phenomena by appealing solely to PP properties, then seemingly any answer to (a) will be *ad hoc*, in which case we will fail to provide a plausible answer to (b). It is unclear how the Non-reductionist can satisfy the burden of proof here and show how we ought to respond to (a) and (b) non-arbitrarily.

Until this burden of proof is met, Non-reductionism fails to offer any good reason to grant the existence of the self as a *sui generis* entity. Reductionism also seems unsatisfactory. Because these views apparently exhaustively explain the relation between self (as intuitively understood) and psychophysical complex, Nāgārjuna concludes that the self does not exist. Granting this argument's plausibility, we should conclude, with Nāgārjuna, that the self, as intuitively understood, does not exist. Whether Nāgārjuna's argument is conclusive, I cannot determine here, due to this paper's scope. Instead, I describe the Buddhist deflationary account of selfhood.

4. The Buddhist View of Selfhood

Conventional Functional Persons

Buddhists (and I) distinguish between 'self' and 'person.' A *person* consists of some conventional label that we apply to a set of psychophysical complexes that are causally continuous over time, while the *self* (as intuitively understood) is the essential feature in virtue of which a person has diachronic identity.[43] Although Buddhists reject that the self exists, they do

[43] Siderits, *Buddhism as philosophy: An introduction*, 32.

not reject that the *person* exists.[44] More specifically, Buddhists deny that *either* selves or persons comprise the ontology of the world because, Buddhists suggest, ultimately our ontology consists entirely of impersonal psychophysical processes.[45] Nevertheless, Buddhists grant that there is a sense in which the concept of a *person*—but *not* our *intuitive* concept of selfhood—may be (coherently) constructed or built out of our ontological concepts of psychophysical processes. Consequently, Buddhists take *persons* to be 'conceptual fictions' that we conceptually construct out of our more fundamental ontological concepts of impersonal psychophysical processes.[46,47]

Altogether, Buddhists hold that we apply our concept of personhood as a conceptual fiction or logical construction to socially embedded organisms as such organisms consist of sets of causally continuous psychophysical processes.[48] Overall, the Buddhist account of *personhood* provides a positive response to the question "What am I?" so it counts as a philosophically sophisticated, non-intuitive account of *selfhood*. In other words, the Buddhist concept of

[44] Collins, *Selfless persons: Imagery and thought in Theravāda Buddhism*, 79; Siderits, *How things are: An introduction to Buddhist metaphysics*, 18.

[45] Ibid.; Siderits, *Buddhism as philosophy: An introduction*; Siderits, Buddhist reductionism; Sauchelli, "Buddhist reductionism, fictionalism about the self, and Buddhist fictionalism."

[46] See Collins, *Selfless persons: Imagery and thought in Theravāda Buddhism*, 103–10; Rahula, *What the Buddha taught*, 51–66; Siderits, "Buddhist reductionism"; Siderits, *Buddhism as philosophy: An introduction*, 22–24, 26, fn. 10.

[47] Chisholm (1976) makes a useful distinction that is relevant here between *entia per se* and *entia per alio*. In contrast to *entia per se*, *entia per alio* are "ontological parasites that derive their properties from other things," and which "never [are] or [have] anything on [their] own," but "[are] what [they are] in virtue of the nature of something other than [themselves]" (p. 104). Consequently, *entia per alio*, unlike *entia per se*, do not exist in the 'strict and philosophical sense,' but only in a 'loose and popular sense.' I think that Buddhists would grant Chisholm's distinction between things that exist in a 'strict and philosophical sense' (i.e., *entia per se*) and things that exist in a 'loose and popular sense' (i.e., *entia per alio*). Given this distinction—unlike Chisholm—Buddhists would suggest that *persons* are *entia per alio*, not *entia per se*.

[48] Garfield, *Losing ourselves: Learning to live without a self*, 5; Richards, "Conceptions of the self in Wittgenstein, Hume, and Buddhism: An analysis and comparison," 51; Sauchelli, "Buddhist reductionism, fictionalism about the self, and Buddhist fictionalism"; Siderits, *How things are: An introduction to Buddhist metaphysics*, 29–46.

personhood is itself a non-intuitive concept of selfhood. To avoid confusion, I use 'person' here to refer to the *positive* Buddhist concept of selfhood, given that I have been using 'self' to refer to our intuitive self-concept.

The concept of personhood on the Buddhist view is importantly different from the intuitive concept of selfhood, in more ways than I can describe here.[49] Most importantly, a person is essentially embodied: she is not ontologically independent of some psychophysical complex, nor does she meaningfully constitute an 'owner' of this complex. Persons are also embedded within the world as a kind of natural phenomenon and, consequently, are causally interdependent with other natural phenomena.[50] Finally, because an organism must fulfill some (proper) social role to be a person[51], I suggest that personhood, unlike our intuitive concept of selfhood, is what I call a *conventional functional concept*. Altogether, according to the Buddhist view, persons lack the essential features that characterize our intuitive concept of selfhood: priority, unity, duality, and agency.

Since personhood constitutes a conventional functional concept, persons are individuated based on what they *do* or the *roles* they play[52], where these roles are grounded in social conventions.[53] Moreover, the conventionally grounded functional property of being a (particular)

[49] See Collins, *Selfless persons: Imagery and thought in Theravāda Buddhism*, 71–78; Sauchelli, "Buddhist reductionism, fictionalism about the self, and Buddhist fictionalism"; Siderits, "Buddhist reductionism"; Siderits, *Buddhism as philosophy: An introduction*, 56–58; Siderits, *How things are: An introduction to Buddhist metaphysics*, 29–46.
[50] Garfield, *Losing ourselves: Learning to live without a self*, 21.
[51] Ibid, 42.
[52] Carlisle, "Becoming and un-becoming: The theory and practice of *anatta*," 77.
[53] See Siderits, *How things are: An introduction to Buddhist metaphysics*, 29–46.

person can be implemented by different entities at different times because different entities may play the same conventional functional role at different times. Since the concept of 'person' is conventionally grounded, persons possess diachronic identity in a *conventional* sense: Persons are akin to characters in a play, who persist across contexts and times while played by different actors.[54] Consequently, changes in psychophysical facts do not imply the existence of different persons over time. If different psychophysical complexes play the same conventional functional role at times t_1 and t_2, then we can meaningfully say that the same person exists at both t_1 and t_2, even though this role is being played at different times by different psychophysical complexes.

As a conventional functional concept, personhood is a folk psychological concept. Andrews (2020) describes folk psychology as consisting in "seeing others as intentional agents with their own traits and goals who are embedded in a community of others."[55] We employ folk psychology as a kind of theoretical framework to explain others' actions and behaviors in terms of the desires and beliefs that we attribute to them using the same theory.[56] Folk psychological explanations are *functional* in nature because they treat desires and beliefs as inputs productive of actions as outputs.[57] The conventional functional concept of personhood is a folk psychological concept because we use it when we engage in folk psychological explanations of why individuals behave as they do.

[54] Garfield, *Losing ourselves: Learning to live without a self*, 37–43.
[55] Andrews, *The animal mind: An introduction to the philosophy of animal cognition*, 31.
[56] Ibid.
[57] Ibid, 33.

A Buddhist Operational Definition of Personhood

If personhood is a conventional functional, folk psychological concept, then members of other species may plausibly satisfy the criteria for this concept and count as being *persons*. Unlike our intuitive concept of selfhood, there is no *conceptual* reason to deny that other animals may count as being persons because, like human beings, at least some other animals are embodied beings who are causally embedded in the natural world and fulfill certain kinds of social roles. By attributing personhood to other animals, we do not commit a category error.[58]

Using the Buddhist account above, let us introduce the notion of a 'Personal Description,' abbreviated PD. Plausibly, an organism is a person if and only if they satisfy some PD. A PD specifies some exhaustive set of behavioral and psychological dispositions, habits, and social roles (all indexed to time). Ideally, a PD specifies a complete functional description of what it means to be a *particular* person. As such, a fully specified PD must describe the behavioral, social, and psychological characteristics of a particular person so completely and uniquely that it is very unlikely that this PD would be satisfied by more than one organism at a moment in time. Let us define personhood thus:

> Some organism *x* is a person if and only if *x* implements some Personal Description (PD), which is an exhaustively complete functional description consisting of some set of social roles, behavioral dispositions, psychological dispositions, emotional dispositions, etc.

[58] That is, we do not make a category error by attributing personhood to other animals, even if as a matter of empirical fact no other animals are persons. This is because it is conceivable that at least some other animals satisfy the necessary and sufficient conditions for being persons. In contrast, it is not conceivable that other animals satisfy the necessary and sufficient conditions for being selves, according to our intuitive concept of selfhood.

Consequently, if we define personhood in terms of the implementation of a PD, which specifies a complete functional description of what it means to be a particular person, then we thereby specify the necessary and sufficient conditions for being a particular person. An organism who implements a PD over time will display psychological and behavioral continuity over time. Thus, this definition allows us to capture diachronic personal identity in the conventional sense, described above, of playing the same "character" over time.[59] Altogether, like the Buddhist concept of personhood, a PD is a *conventional functional* kind of description.

We can construct a tentative operational definition on the basis of this definition of personhood. Since a PD is a functional kind of description of a person, and since operational definitions utilize functional descriptions, we can use the content of a PD in our operational definition, where the content of a PD concerns psychological and behavioral continuity over time. Therefore, determining whether an organism is a person on the basis of the operationalization of our concept of personhood will depend on that organism's behavioral and psychological continuity over time.

Another important feature to consider when ascribing personhood to an organism *x* is how *other* organisms engage with *x*. We engage with persons differently depending on which sets of behavioral and psychological dispositions we attribute to them in our interactions with them. Similarly, determining when to ascribe personhood to an organism *x* on the basis of our operationalization of this concept would benefit from considering the *reactive* behavioral

[59] See Garfield, *Losing ourselves: Learning to live without a self*, 37–43.

dispositions of *x's* conspecifics (or non-conspecifics).

As a tentative operational definition of personhood, let us say that:

> An organism *x* is a person if and only if (1) *x* exhibits a unique pattern of behavioral and psychological traits (as specified by some PD) over a significant period of time and across a diverse set of contexts and roles; and (2) *x* engages with conspecifics (or non-conspecifics) who exhibit consistent patterns of behavioral and psychological traits in their interactions with *x*.

Specifying what a "significant period of time" or "a diverse set of contexts and roles" consists in requires further analysis that cannot be completed here due to space. Further analysis of the nature of 'role' is also pertinent. 'Uniqueness' here *tentatively* consists in an organism's implementing some set of behavioral and psychological traits specified on some ideally exhaustively described PD. Furthermore, if conspecifics (or non-conspecifics) interact with an organism *x* by exhibiting consistent patterns of behavioral or psychological traits, then these conspecifics (or non-conspecifics) are likely tracking the behavioral and psychological traits uniquely instantiated by *x*. Consequently, as I point out above, identifying these consistent patterns of interaction may be pertinent to identifying the personhood of animals.

Using this operational definition, I now provide some empirical evidence for the claim that some animals may be persons in the same sense in which human beings are.

5. Evidence for the Personhood of Non-Human Animals

Empirical evidence suggests that members of some species may plausibly satisfy the proposed definition for personhood. As Bekoff (2003) notes, members of various species—including chimps, rhesus monkeys, wolves, crows, bears, and even sweat bees and ants—each interact and communicate in various contexts. The ability to consistently interact across various contexts may require behavioral and psychological continuity to undergird and facilitate communication. Consequently, members of some, or all, of these species likely display behavioral and psychological continuity in their communicative interactions across contexts and roles. Ergo, members of these species count as *prima facie* candidates for personhood.

Furthermore, ethological research suggests that members of various species exhibit personality traits, including great tits, octopuses, dogs, and orangutans.[60] Additional research suggests that even some *insects*, including bees and crickets, may display personality traits.[61] If an organism exhibits personality traits, then *ipso facto* that organism has behavioral and psychological continuity across time. Thus, species whose members demonstrate variable personality traits *ipso facto* count as species consisting of *prima facie* candidates for personhood.

Further research also suggests that members of certain species, including chimpanzees and orangutans, apparently understand personality differences among conspecifics.[62] That members of these species track personality differences illustrates that they track behavioral and psychological continuity among conspecifics. This suggests both that the conspecifics whose

[60] See Amy et. al., "Effects of personality on territory defense in communication networks: A playback experiment with radio-tagged great tits"; Mather & Anderson, "Personalities of octopuses (octopus rubescens)"; Gosling & John, "Personality in non-human animals"; Weiss et. al., "Personality and subjective well-being in orangutans (Pongo Pygmaeus and Pongo Abelli)"; Freeman & Gosling, "Personality in non-human primates: A review and evaluation of past research."

[61] Walton & Toth, "Variation in individual worker honey bee behavior shows hallmarks of personality"; Gosling, "Personality in non-human animals."

[62] Subiaul et. al., "Do chimpanzees learn reputation by observation? Evidence from direct and indirect experience with generous and selfish strangers"; Herrmann et. al., "Direct and indirect reputation formation in nonhuman great apes (Pan Paniscus, Pan Troglodytes, Gorilla Gorilla, Pongo Pygmaeus) and human children (Homo Sapiens)."

personality traits are perceived may be persons and, if this claim is justified, that the perception or identification of personhood may not be unique to humans.

Finally, Andrews (2020) provides evidence for the existence of social norms among certain species, such as chimpanzees.[63] She operationalizes the concept of social norms thus:[64]

> A social norm is to be identified by the existence of three elements: (a) There is a pattern of behavior demonstrated by community members; (b) individuals are free to conform to the pattern of behavior or not (the behavior is voluntary); and (c) individuals expect that community members will also conform, and will sanction those who do not conform.

Since this definition builds behavioral continuity into it, members of any species satisfying this definition *ipso facto* count as being *prima facie* candidates for personhood. Also, this definition requires that individuals expect community members to conform to certain patterns of behavior. Organisms with these expectations likely track community members' unique sets of psychological and behavioral traits. If so, these organisms might satisfy the second clause of the operational definition of personhood. Importantly, satisfying this clause (and Andrews's definition for social norms) does not require the capacity to mindread.[65] All that is required is that conspecifics can *behaviorally* track an organism's unique set of psychological and behavioral traits.

Altogether, using this operational definition of personhood based on the Buddhist view, we have preliminary reasons to suspect that further empirical evidence will favor attributing

[63] Andrews, *The animal mind: An introduction to the philosophy of animal cognition*, 220–21.
[64] Ibid, 218.
[65] *Mindreading* consists in the ability to infer others' mental states based on observable behavioral cues. See Lurz, "Animal mindreading: The problem and how it can be solved," 229.

personhood to members of other species. The Buddhist account of personhood serves as a deflationary, but philosophically sophisticated account of selfhood. Thus, if we have preliminary evidence for some animals being 'persons' on the Buddhist view, then these animals may be candidates for selfhood in a special, philosophically sophisticated sense (not in our intuitive, pre-theoretic sense). Since we can appeal to the conceptual scheme constituting the Buddhist view of personhood to plausibly ascribe full-fledged selfhood to other animals, conceptual reasons need not bar animals from being 'selves' in the same sense in which humans are.[66]

In conclusion, if we accept the Buddhist view of personhood over our intuitive view of selfhood, then whether other animals are full-fledged selves depends on what the empirical evidence dictates. Apparently, the empirical evidence suggests that members of some species *do* in fact possess the necessary and sufficient psychological and behavioral traits for consummate selfhood in the sense of Buddhist personhood.

6. Conclusion

I have argued that if we draw upon the Buddhist deflationary account of selfhood to develop an operational definition of personhood, then we can attribute selfhood (in the sense of Buddhist personhood) to some animals. However, I have not argued that the Buddhist view is conclusive. I have adduced this view only to argue that conceptual reasons need not bar us from ascribing the

[66] In other words, since the Buddhist concept of personhood is itself a non-intuitive concept of selfhood, and since it is conceivable that we can subsume the members of at least some other species under the extension of this concept of personhood, it follows that it is conceivable that there is *some* concept of selfhood such that subsuming other animals under the extension of this concept does not entail a category error.

same kind of selfhood to other animals as we ascribe to ourselves. If the Buddhist deflationary view is plausible, and if the tentative suggestions of the empirical evidence are correct, then we have one less reason to think that the difference between human and non-human animals consists in a difference of *kind*.

Works Cited

Amy, M., Sprau P., de Goede, P., & Naguib, M. (2010). "Effects of personality on territory defense in communication networks: A playback experiment with radio-tagged great tits." *Proceedings of the Royal Society B: Biological sciences, 277*(1700), 3685–3692.

Andrews, K. (2020). *The animal mind: An introduction to the philosophy of animal cognition.* Routledge.

Bekoff, M. (2003). "Considering animals—not "higher" primates: Consciousness and self in animals: Some reflections." *Zygon, 38*(2), 229–45. 0591-2385

Carlisle, C. (2006). "Becoming and un-becoming: The theory and practice of *anatta*." *Contemporary Buddhism, 7*(1), 75–89. https://doi.org/10.1080/14639940600878034

Chisholm, R. (1976). *Person and object: A metaphysical study*. Routledge.

Collins, S. (1982). *Selfless persons: Imagery and thought in Theravāda Buddhism*. Cambridge.

Dornbach, M. (2023). "Animal selfhood and affectivity in Helmuth Plessner's philosophical biology." *Philosophical forum, 54*(4), 201–30. DOI: 10.1111/phil.12340

Freeman, H. D., & Gosling, S. D. (2010). "Personality in non-human primates: A review and evaluation of past research." *American journal of primatology, 72*(8), 653–71.

Ganeri, J. (2012). *The self: Naturalism, consciousness, and the first-person stance.* Oxford.

Garfield, J. (2022). *Losing ourselves: Learning to live without a self*. Princeton.

Gosling, S. D., & John, O. P. (1998). "Personality dimensions in dogs, cats, and hyenas." *Annual meeting of the American Psychological Society*, 1998.

Gosling, S. D. (2008). “Personality in non-human animals.” *Social and personality psychology compass, 2*(2), 985–1001. 10.1111/j.1751-9004.2008.00087.x

Herrmann, E., Keupp, S., Hare, B., Vaish, A., & Tomasello, M. (2013). “Direct and indirect reputation formation in nonhuman great apes (Pan Paniscus, Pan Troglodytes, Gorilla Gorilla, Pongo Pygmaeus) and human children (Homo Sapiens).” *Journal of comparative psychology, 127*(1), 63–75.

Lurz, R. (2018). “Animal mindreading: The problem and how it can be solved.” In K. Andrews & J. Beck (Eds.), *The Routledge handbook of philosophy of animal minds* (pp. 229–37). Routledge.

Mather, J. A., & Anderson, R. C. (1993). “Personalities of octopuses (octopus rubescens).” *Journal of comparative psychology, 107*(3), 336–40.

Nāgārjuna. (1995). *Mulamadhyamakakarika* [*The fundamental wisdom of the middle way*]. Translated by J. Garfield. Oxford.

Rahula, W. (1959). *What the Buddha taught.* New York: Grove Press.

Richards, G. (1978). “Conceptions of the self in Wittgenstein, Hume, and Buddhism: An analysis and comparison.” *The monist, 61*(1), 42–55. https://www.jstor.org/stable/27902511

Saidel, E. (2018). “On psychological explanations and self concepts (in some animals).” In K. Andrews & J. Beck (Eds.), *The Routledge handbook of philosophy of animal minds* (pp. 131–41). Routledge.

Sauchelli, A. (2016). "Buddhist reductionism, fictionalism about the self, and Buddhist fictionalism." *Philosophy east and west, 66*(4), 1273–1291. https://doi.org/10.1353/pew.2016.0091

Siderits, M. (1997). "Buddhist reductionism." *Philosophy east and west, 47*(4), 455–478. https://www.jstor.org/stable/1400298

Siderits, M. (2007). *Buddhism as philosophy: An introduction*. Hackett Publishing.

Siderits, M. (2015). *Personal identity and Buddhist philosophy: Empty persons* (2nd edition). Routledge.

Siderits, M. (2022). *How things are: An introduction to Buddhist metaphysics.* Oxford.

Subiaul, F., Vonk, J., Okamoto-Barth, S., & Barth, J. (2008). "Do chimpanzees learn reputation by observation? Evidence from direct and indirect experience with generous and selfish strangers." *Animal cognition, 11*(4), 611–23.

Walton, A., & Toth, A. L. (2016). "Variation in individual worker honey bee behavior shows hallmarks of personality." *Behavioral ecology and sociology, 70*(7), 999–1010. DOI 10.1007/s00265-016-2084-4

Weiss, A., King, J. E., & Perkins, L. (2006). "Personality and subjective well-being in orangutans (Pongo Pygmaeus and Pongo Abelli)." *Journal of personality and social psychology, 90*(3), 501–11.

Can Turing Machines Possess Intrinsic Intentionality?

Zhen Wang

This paper explores the question of whether Turing machines, particularly artificial intelligence (AI) systems, can exhibit intrinsic intentionality — defined as the capacity to interpret internal processes and generate meaningful outputs. This paper then discusses Searle's Chinese Room Argument (1980), which challenges the possibility of machines' intrinsic intentionality, as well as the syntactic theory that suggests otherwise. This theory suggests that internalized syntactic processes suffice for creating intrinsic intentionality. Rapaport used Helen Keller's experience to illustrate how the internalization of symbols may create intrinsic intentionality (2007) . Finally, this paper raises objections to syntactic semantics as a solution to Turing Machines to acquire intrinsic intentionality. It argues that AI symbols can only be about intrinsically meaningless tokens without phenomenon experience. Drawing on Jackson's Knowledge Argument (1982), the paper contends that intrinsic intentionality requires a mental process to be about a phenomenal experience.

1. Introduction

For humans, our mental activities have meaning. To say that all raccoons are mammals is not merely a logical proposition that all *A*s are *B*. For us, a raccoon means a bandit-looking furry creature with four limbs and various other characteristics. We can visualize one with our mind's eye and imagine how it moves or sounds. Computers are Turing machines that manipulate inputs based on sets of instructions. An artificial intelligence system may contain a class called mammal which has a subclass called raccoon in its storage. But does a raccoon mean anything to this system when it processes a raccoon? I will first discuss Searle's Chinese Room Argument as a negative answer to this question. Then, I will present and evaluate the theory of syntactic semantics which argues that internalized syntactic processes are meaningful on their own. Finally, I will argue against the syntactic semantics theory by arguing that the grasp of meaning requires intrinsic intentionality, which requires phenomenon consciousness.

2. Two Types of Intentionalities

In keeping with influential works in Philosophy of Mind, I use the term intentionality to mean "the power of a process to be directed at or about certain things like objects, properties, and states of affairs."[67] There are two types of intentionality: original intentionality and derivative intentionality.[68] A book, for example, can refer to many objects or concepts through its texts. However, it only does so when a reader interprets it. So, the book only has derivative intentionality that affords its interpretability. Such intentionality was given by the author of the book and reconstructed by its readers. Original intentionality is the capability of delegating representations to objects and interpreting objects from representations. Therefore, original intentionality exists only in the interpreters of the book. For the purpose of this essay, I will refer to original intentionality as intrinsic intentionality. This is because the word "original" may carry a connotation of authorship in the legal sense. An interpreter of words in a book possesses original intentionality not because they are the first to delegate certain meaning to the words, but

[67] Searle, 1980; Haugeland, 1990; etc. Dietrich et al. 2021, p. 93

[68] Haugeland, 1990

because they are capable of delegating *any* meaning to them.

3. The Chinese Room Argument and Intentionality

The problem of machines and meaning is not about derivative intentionality. The outputs of machines like a calculator or a large language model (LLM) can usually sustain human interpretation. This is because their symbols can be translated into a human language, and their syntax can be defined to only allow interpretable outputs. If you take care of the syntax, the *derivative intentionality* will take care of itself. However, it is far from clear whether a machine can possess intrinsic intentionality — the power to interpret its internal processes and produce sensible output that is also meaningful to itself. This is an apparent feature of human cognitive systems. We can *interpret* what we think (our internal processes), what we say, and even much of what others say. A famous argument against the possibility of artificial intelligence (AI) having intrinsic intentionality is the Chinese Room Argument proposed by Searle (1980). He wondered whether the human mind works like a Turing machine, a purely formal system. He concludes that if we work like that, we would not be able to even interpret our own languages.

Suppose you are locked inside a room with an input slot and an output slot. The input you receive is written in a language completely strange to you. There is a handbook that outlines how you should respond upon encountering any kind of input. So, being a good rule follower, you produce correct responses and insert them into the output slot. To an outsider who understands the strange language, it is as if the room has a native speaker of that language.

Searle notes that no matter how good you are at manipulating the inputs to produce the outputs, the language means nothing to you. Searle notes that in the Strange Language Room, you behave just like a computer processor. The handbook is like a program written by intelligent programmers. While you do not understand that strange language, the book enables you to pretend to understand. Therefore, if an otherwise-intentional being like yourself cannot derive intentionality from formal syntactic operations, there is no reason to believe a computer processor can. What gives us intrinsic intentionality must not be formal syntactic manipulation.

For the machine to possess intrinsic intentionality, it needs to be able to interpret its own processes and figure out what they are *about*. Searle argues that human brains have "proper causal powers" to possess intrinsic intentionality. Searle does not argue that our intentionality must represent the outside world. Those proper causal powers refer to the physical-chemical processes and the biological structure of an organism's brain.[69] This implies that a brain-in-a-vat would possess intrinsic intentionality, whereas a silicon-based robot that can act entirely indistinguishable from humans never could. However, Searle makes no argument defending how biological processes, *but not* electronic processes, can give rise to intrinsic intentionality. If this claim is not taken for granted, then another compelling theory of intrinsic intentionality should be considered.

4. Syntactic Semantics

In response to Searle (1980), Dietrich et al. (2021) discuss the syntactic semantics theory of

[69] Searle 1980, p. 442.

intentionality. Proponents of syntactic semantics believe that a formal system is sufficient to generate intrinsic intentionality. Rapaport (2007) uses the life story of Helen Keller to argue that intentionality arises when all semantics are properly internalized. Helen Keller was both blind and deaf since childhood, yet she could learn to communicate using finger gestures, signs, and eventually English. Rapaport argues that Keller had been living in a version of Searle's Strange Language Room for almost her entire life. The following quote from Keller's autobiography suggests that she manipulated the English symbols based on syntactic rules: "I did not know that I was spelling a word or even that words existed; I was simply making my fingers go in monkey-like imitation."[70] As she mastered the syntax, it was obvious that she does understand English, and English *means* something to her. Dietrich et al. (2021) summarize that the key to syntactic semantics is the internalization of external symbols. Once they are appropriately internalized by the agent, the symbols are intrinsically intentional.

Her example seems to suggest that formal syntactic manipulation can be sufficient for intrinsic intentionality. Computers are good at syntactic manipulation, so perhaps they can possess intrinsic intentionality as well. Under Rapaport's syntactic semantics theory, symbols can be said to be *about* each other via a process called variable binding. This process lets a variable name refer to an entity. Variables and objects can be defined in terms of each other and constitute each other. The AI interprets a variable by following its references. For example,

[70] Keller, 1905, p. 35, cited in Rapaport, 2007, p. 395.

suppose an AI system with cameras detected a raccoon sleeping on the grass.[71] The object recognition algorithms determined that the object was a raccoon. So, the AI instantiated a Raccoon in its environmental model with the following *fields* (encapsulated information in object-oriented programming languages):

Raccoon #3942	
Class	Animal
Sub-class	Raccoon
Colour	R: 23 G: 21 B: 27
Distance	5
Ground Velocity	2

So, Raccoon #3942 referred to the combination of all its properties/fields. Variables such as “distance” and “ground velocity" referred to numbers five and two. As the robot approached, it startled the raccoon who increased its velocity away from the robot. So, the robot retrieved those variables and incremented them as such:

[71] cf. Dietrich et al., 2021, pp. 98-9.

Raccoon #3942	
Class	Animal
Sub-class	Raccoon
Colour	R: 23 G: 21 B: 27
Distance	Distance + 4
Ground Velocity	Ground Velocity + 5

The AI being able to follow references is a sign of interpretation according to the syntactic semantics theory. The raccoon can also exist in relation to other objects in the robot's environmental model. For instance, a new field in Raccoon #3942 called Surface can indicate the surface on which it stands. Surface can be bound to a grass chunk. The grass chunk can also have a field that is bound to Raccoon #3942. If the environment model is set up properly, the AI system can simulate interactions between objects and run counterfactual scenarios. This does seem to approach the power of human intentionality about other objects. Note that AI may behaviorally simulate intentional beings like humans. The physics simulation above can afford it to perform some goal directed actions. However, the question of whether variable binding captures all that is required for having intrinsic intentionality still remains to be open.

5. Meaningless Symbols Do Not Produce Meaning

On the table representations of Raccoon #3942, I intentionally (no pun intended) included a field called Color to raise suspicion about the syntactic semantics theory. The AI system represented colors using the intensity of primary colors: red, green and blue. The raccoon's color may fall under gray to a sighted human. But what is the Color field *about* to the AI system? It seems that it is really about a collection of three integers: R, G, and B. Then what does each of them mean? A knowledgeable computer scientist may program all we know about color science into the AI system. However, does that give it any idea about what red, green, or blue means? This scenario is analogous to Mary's (the color scientist) situation in Jackson's Knowledge Argument for qualia (1982). I believe if we *somehow* programmed the phenomenal experience of seeing colors into the AI system, it would learn something new. Without being able to experience any color, the AI's color field cannot be color.

An objection is that the syntactic AI system can experience colors via the camera connected. The experience involved the sensor registering lights of different frequencies, the processor writing data into the memory, and so on. So, there is no need to *somehow* program the phenomenal experience because the system could already experience colors. My response is that attributing phenomenal experience to camera sensors and processors risks anthropomorphizing mechanical processes. There are three premises for my response.

(1): Phenomenal experience requires levels of dynamical emergence.

(2): The light-sensitive material in a camera's sensor does not sustain the levels of emergence.

(3): AI systems designed for accomplishing computation do not sustain the levels of emergence.

The first premise is based on works of Terrence Deacon on biological anthropology and neuroscience. Deacon (2013) argues that phenomenal consciousness requires three levels of emergence from thermodynamic (homeodynamic) processes to morphodynamic, teleodynamic, and higher-order teleodynamic processes.

> ... this second-order teleodynamics is analogous to the way that the teleodynamics of interacting organisms within an ecosystem can contribute to higher-order population dynamics, including equilibrating (homeodynamic) and self-organizing (morphodynamic) population effects… the tendency for population-level morphodynamic processes to emerge in the recursive flow of signals within a vast extended network of interconnected neurons is critical to the generation of mental experience … This tangled hierarchy of causality is responsible for the special higher-order sentient properties (e.g., subjective experience) that brains are capable of producing, which their components (neurons) are not.[72]

He argues that sentience is the result of organisms (perhaps not exclusive to biological ones) engaging in self-creative and self-bounding tendencies. I argue that AIs that work like a Turing Machine function only at the thermodynamic level, and are neither self-creative nor self-bounding. The same can be said about the camera sensor. Therefore, I take (2) and (3) as true. If all my premises are true, it follows that:

[72] Deacon, 2013, p. 510.

> (4): the phenomenal experience of color cannot come into existence by connecting a light sensor and processors without additional emergent processes.

If we "interrogate" the AI system for the meaning of a color, it can only respond with other ungrounded symbols. Note that Searle (1980) would not even consider those as symbols because they are not interpretable for machines (p. 422). For a symbolic AI (in contrast to artificial neural networks), I grant Rapaport (2007) that a symbol can refer to the symbol(s) that it was bound to. For artificial neural networks (ANNs), there are no longer distinct high-level symbols interpretable to humans. Their operations consist of layers of threshold logic units (TLUs) and store information in their connection weights.[73] They are trained with input data and using algorithms like error backpropagation to modify thresholds in TLUs to produce desirable outputs.[74] This means that they are Turing Machines that perform syntactic operations on their inputs. However, what do the syntactic operations mean? The compiler of a program translates executable high-level computer instructions into low-level instructions. Eventually, the codes are translated into machine-readable binary instructions. Binary instructions are actualized in the silicon as different voltages in wires and logical gates. Nowhere in these processes could a phenomenal experience seem to emerge.

If a symbol is not fundamentally about a phenomenal experience, what could it be about? My answer is meaningless tokens. For a person who has never experienced the color red, there

[73] Kruse et al., 2013, p. 15

[74] Ibid, 34, 66.

is no amount of mental gymnastics they can do to make $\langle R\text{: } 255,\ G\text{: } 0,\ B\text{: } 0\rangle$ about this color 🟥 (a red colored square). For a person who can see red, they can try to imagine a color that is outside of the human's visible spectrum. We can think about the light's (or electromagnetic radiation) physical or thermal properties because they can translate into our experience, but we can never think of that color.

I propose that a mental process is intrinsically intentional *if and only if* it is about a phenomenal experience. A problem with Rapaport's (2007) analogy that Helen Keller lived in a Strange Language Room is that she lived *the human experience.* She experienced emotions and sensations. Her concepts of water, cake, coldness, and textures of objects were all grounded in the sensations that they cause. This is vastly different from a purely symbol manipulator such as our AI friend above. All its symbols only refer to other symbols, whereas Keller's finger plays could refer to phenomenal experiences.

A corollary of this proposal is that not all human mental processes are intentional. Processes about purely syntactic constructs are only derivatively intentional. For example, when I *only* think of the number two, it does not refer to any phenomenon. It could refer to the successor of one or the predecessor of three in the domain of integers, but those references are only syntactical because both one and three are also mere syntactic constructs. In contrast, to think of two apples is about the phenomenon of them; a combination of their colors, smell, taste, texture, etc. Of course, I can think of "two apples" as an abstract symbol. This would make the thought *only* derivatively intentional. The act of interpreting the symbol can ground it to something phenomenal and thus make it intrinsically intentional.

6. Conclusion

To build an AI that thinks like humans, intrinsic intentionality is an important feature that needs to be included. The human mind is intrinsically intentional because we can interpret what our own mental activities are about. Searle's Chinese Room Argument (1980) demonstrates that no amount of syntactic manipulation can give rise to intrinsic intentionality. He further argues that only biological brains are capable of generating intrinsic intentionality, but he does not give sufficient evidence for this claim. Therefore, it seems promising that the syntactic semantics theory could tackle the challenge posed by Searle (1980). Rapaport (2007) proposes that appropriately internalizing symbols into a system is sufficient to create intrinsic intentionality, regardless of human brains or Turing machines. He suggests that Helen Keller learned a human language via a similar process. I argue that Rapaport understated the importance of Keller's phenomenal experience as a human being. It was the human experience that provided something to ground her symbols onto. I propose that a process is intrinsically intentional if and only if it is about a phenomenal experience. I am not convinced that any Turing machine-based AI has phenomenal experience. Thus, they are not intrinsically intentional. However, I do not exclude the possibility of AI acquiring phenomenal experience someday. How AI might gain phenomenal experience is an important question for future research.

Works Cited

Deacon, T. W. (2013). "Incomplete Nature: How Mind Emerged from Matter." W. W. Norton.

Dietrich, E., Fields, C., Sullins, J.P., van Heuveln, B., & Zebrowski, R. (2021). "The Strange Case of the Missing Meaning: Can Computers Think About Things?". In Great Philosophical Objections to Artificial Intelligence: The History and Legacy of the AI Wars (pp. 87–134). London: Bloomsbury Academic. Retrieved November 30, 2023, from http://dx.doi.org/10.5040/9781474257084.ch-005

Haugeland, J. (199 0). "The Intentionality All-Stars." *Philosophical Perspectives*, *4*, 383–427. https://doi.org/10.2307/2214199

Keller, H. (1905). *The story of my life*. Garden City, NY: Doubleday (1954).

Jackson, F. (1982). "Epiphenomenal Qualia." The Philosophical Quarterly (1950-), 32(127), 127–136. https://doi.org/10.2307/2960077

Kruse, R., Borgelt, C., Klawonn, F., Moewes, C., Steinbrecher, M., &#38; Held, P. (2013). "Threshold logic units." In Texts in Computer Science (pp. 15–35). Springer London. http://dx.doi.org/10.1007/978-1-4471-5013-8_3

Rapaport, W. J. (2007). “How Helen Keller used syntactic semantics to escape from a Chinese Room.” Minds and Machines, 16(4), 381–436. https://doi.org/10.1007/s11023-007-9054-6

Searle, J. (1980). “Minds, brains, and programs.” Behavioral and Brain Sciences, 3(3), 417-424. doi:10.1017/S0140525X00005756

Care Bots & The Issue of Deception

Ila Kacker

The aging population in America is growing faster now than ever. However, we lack the proper infrastructure and resources to care for them adequately. Those involved in the field of elder care are experimenting with solutions to this problem. One of the most pressing solutions is the use of artificial intelligence, namely care bots. Care bots are a specific type of technology that aims at providing physical and emotional support for the vulnerable elderly population. While the practical benefits of care bots are evident, the ethical implications relating to social isolation, paternalism, and deception must also be considered before they can be implemented as caregivers. With a specific focus on the issue of deception, I will demonstrate that certain types of care bots, such as those that simulate a reciprocal relationship between the bot and the care receiver, are inherently deceptive and immoral. However, other types of care bots, such as nurse bots, may be ethical as they do not attempt to simulate a reciprocal relationship, and they act in a manner consistent with benevolence rather than care.

1. An Aging America

The aging population in America is greater now than ever, with the population of those 65 and older growing almost five times faster than the rest of the population over the past 100 years. To put this number in perspective, in 1920, less than 1 in 20 people were over the age of 65, now about 1 in 6 people are.[75] With the rise of modern medicine and the improvement of health outcomes, the elderly population is living longer than ever. However, it is important to

[75] Bureau, U. C., *U.S. Older Population Grew From 2010 to 2020 at Fastest Rate Since 1880 to 1890.*

acknowledge that American infrastructure is not built to accommodate the influx of this demographic. Especially considering the extensive care and supervision many within this population may need, it is easy to see how nursing homes, medical personnel, and caregivers may become stressed and overwhelmed due to the increased population size. Many researchers have turned to the idea of artificial intelligence to combat this issue. Artificial intelligence, which is beginning to be more frequently used in healthcare contexts, specifically the robot technology coined "the care bot," is one of the most promising options to revolutionize elder care.

2. Care Bot Technology

A general definition of a care bot is a robot that provides care and support for vulnerable people suffering from mental and physical ailments. Many different companies have attempted to develop their own versions of care bots, including the Care-O-bot, Robear, or Actron MentorBotTM, which are all robots that can help the care-receiver with tasks such as those around the household, lifting a patient from their beds to wheelchairs, and reminding patients to take their medication.[76] Beyond the physical assistance, these robots can also provide companionship and comfort. One particular example that has garnered a lot of attention is PARO, the interactive robot made to look like a seal, due to its ability to interact and comfort dementia patients by making sounds and responding to touch.[77] While care bots are most commonly discussed in the context of the elderly and children, they also have the potential to be

[76] Yew, "Trust in and Ethical Design of Carebots: The Case for Ethics of Care," 629–645.
[77] Ibid.

influential in the realm of mental health, addiction, and physical rehabilitation. For the purposes of this paper, however, I am going to focus specifically on the impact of these robots on elder care because I believe it yields interesting philosophical discussion about deception. Furthermore, it is important to note that while these care bots may be used in conjunction with human caregivers as of right now, it seems the hope is that they will eventually be allowed to work autonomously in order to truly alleviate the strain on the elder care system. Therefore, the following arguments assume that care bots are working alone, not as a supplement to human caregivers.

Assuming that these care bots are deemed safe, both in the sense that they will not inflict violence upon the care-receiver, and they will not leak protected health information, they offer numerous practical benefits. For example, the care-receiver can avoid being displaced from their home into a nursing home, keeping their dignity intact. The receivers can also have 24/7 quality care, as the robots will not get fatigued or need any breaks as a human caregiver would. Furthermore, the likelihood for elder abuse is significantly decreased as the robots would be programmed to act in the best interest of the elder. However, while the practical benefits of the implementation of care bots are extremely enticing, it is also essential to discuss the ethical implications of applying this technology to the field of elder care.

3. Ethical Implications of Care Bot Technologies

Prior research has noted that the most fruitful ways care bots can positively aid in elder care consists of assisting them with their daily tasks, monitoring their behavior and health, and providing companionship.[78] These three benefits, however, also have the potential to yield negative ethical implications. A care bot assisting with everyday tasks of the care-receiver may result in that elder having little to no human interaction, leading to social isolation; monitoring their behavior and health may lead to a paternalistic attitude with decreased freedom; and providing companionship may cause deception as it is not possible for a robot to truly care for a human being.

In this section, I will briefly discuss the implications of social isolation and paternalism associated with the use of care bot technology in healthcare, and then I will proceed to focus on the issue of deception with a focus on what it means to care and how moral favorability differs for different types of care bots.

The Issue of Social Isolation

Care bot technology has an immense potential to cater to the physical needs of elders, performing tasks they would normally outsource or need supervision for. The practical benefits of saving time, money, and physical labor are evident. However, the accompanying social isolation is overlooked.[79] Oftentimes, the only human-human social interaction some of these elders have is with caregivers who come to take care of them medically and help with their tasks, while intentionally or unintentionally also providing companionship. This added benefit of companionship has been shown to positively influence health outcomes. Notably, one study

[78] Sharkey, A., and Sharkey, N., "Granny and the robots: Ethical issues in robot care for the elderly," 27–40.
[79] Sharkey, N., and Sharkey, A., "The Eldercare Factory," 282–288.

found that the risk of developing Alzheimer's disease doubled in people who were lonely compared with people who were not lonely.[80] Thus, the issue of social isolation prompts the question of whether a robot-human relationship could ever provide the same companionship as a human-human care relationship. However, it is also important to note that opponents of this view argue that assuming the use of care bot technology leads to social isolation fails to credit the elderly for being able to advocate for their own needs, social or otherwise.[81] This issue lends to a complex and interesting discussion; however, it will not be the focus of this paper.

The Issue of Paternalism

The next issue commonly discussed in the ethics of care bot literature is that of paternalism and the resultant restricted autonomy. Generally, paternalism is defined as the infringement of a person's freedom and autonomy, but more specifically, in the healthcare field, it refers to the confrontation of an individual's autonomy and the well-intentioned social overprotectiveness from others.[82] While the extent of current care bot technology is limited to moving objects out of the way, helping around the house, and providing reminders, with simple extension, it can be easily conceived that they could learn to recognize danger signs and respond to them.[83] These responses include turning off a stove left on or an overflowing bath, which in theory is beneficial. However, acting on any and all danger signs can lead to acting in ways that infringe

[80]Wilson et al., "Loneliness and risk of Alzheimer disease," 234–240.
[81] Coin and Dubljević, "Carebots for eldercare: Technology, ethics, and implications," 553–569.
[82] Fernández-Ballesteros et al., "Paternalism vs. Autonomy: Are They Alternative Types of Formal Care?," 1460.
[83] Sharkey, A., and Sharkey, N., "Granny and the robots: Ethical issues in robot care for the elderly," 27–40.

on the personal freedom of the elder. Furthermore, the care bot would be unable to account for situations where there is no actual danger, but merely a perceived danger.[84] As a result, the care bot would restrict the freedom of the elder and could cause feelings of infantilization. However, it is also important to note that in cases of dementia or cognitive decline, physical interventions may be morally permissible and practically advantageous due to the nature of the elder's illness and their altered mental status. Once again, while this issue yields a very interesting discussion, it will not be the focus of this paper.

The Issue of Deception

The last most commonly discussed ethical issue, which I will spend the rest of the paper discussing, is that of deception associated with the concern that the companionship care bots provide may be mistaken for a genuine, caring relationship. While some care bots have both physical assistive benefits and also provide companionship, others are solely meant to provide companionship, which is an important distinction when thinking about the issue of deception.

The former type of care bot is often referred to as a nurse bot, which is a type of robot meant to emulate some of the functions of live-in nurses as they are able to take vital signs, provide medication reminders, and help with lifting and moving patients. Generally, a nurse bot's main function is to help their elder with assistive tasks, much like the robot in the film *Robot & Frank*. In the film, the Robot helps Frank, who is suffering from dementia, boost his memory via reminders and monitoring his behavior. While the two seem to develop a friendship of sorts

[84] Sharkey, N., and Sharkey, A., "The Eldercare Factory," 282–288.

throughout the film, Robot eventually decommissions himself to help Frank avoid legal ramifications, demonstrating above all, his function was to assist Frank despite also providing companionship.[85] In contrast, consider the aforementioned example of PARO, the fur-covered robotic seal, who can react to being pet and make noises, which is designed in an attempt to act as a companion.[86] By simulating the behaviors of a real animal, there is potential for a bond to be formed between the elder and this care bot, where the nature of the bond is ambiguous since it is very possible that the elder, despite perceiving PARO as an object, may develop feelings of care akin to how one may care for an animal or another person. While certain elders, specifically those with severe dementia, may not understand that PARO is an object, which is a significant moral issue, the more pressing and likely issue is the ability of PARO to invoke feelings of care within the care-receiver based on PARO's behaviors. Overall, these two contrasting examples illustrate the potential confusion when distinguishing human-robot assistance relationships from actual caring relationships.

In the following section, I will argue that care bots such as PARO are guilty of deception because their primary function inherently invokes feelings of care, inevitably resulting in deceit, but other types of care bots like nurse bots are not deceptive because their primary function does not involve caring but rather their actions fall in line with benevolence, the desire to do good to others, not care.

[85] Koistinen, "The (care) robot in science fiction: A monster or a tool for the future?," Article 2.
[86] Sharkey, N., and Sharkey, A., "The Eldercare Factory," 282–88.

In order to care for someone, one must (a) do the right actions to exhibit care for another, and (b) they must do the right actions for the right reasons. The right reasons clause ensures that one is doing something not for their own benefit, but rather for the benefit of whom they are caring for. The first issue that immediately arises when thinking about care bots is that they lack minds, and thus cannot have motivations behind their actions as they are simply programmed to act in a certain manner. Therefore, care bots are unable to truly care, but instead they mimic caring acts involving protecting and helping the care-receiver. However, doing so may be said to be acting in a deceiving manner, as deception is defined as persuading someone that something false is true.[87] Essentially, it can be claimed that by mimicking caring, care bots are deceiving their care receivers by persuading them, through their actions, that they are receiving genuine care. Opponents to this claim can argue that it is pointless to distinguish whether or not the care being received is genuine because either way, the care-receiver is receiving care. However, it is interesting, and perhaps even relevant, to consider the moral implications of mimicking an act.

Those who subscribe to this view believe that mimicking acts of caring is harmful because caring is an inherently valuable virtue. When drawing a comparison to other Aristotelian virtues, the virtue of care is closely tied to that of friendship because care is undoubtedly heavily involved in friendship. Aristotle argues that friendship is inherently valuable as an end itself due to its reciprocal nature where social needs are met through mutual care.[88] Thus, even if robots are

[87] *Deceive*, https://dictionary.cambridge.org/us/dictionary/english/deceive

[88] Elder, "False friends and false coinage: A tool for navigating the ethics of sociable robots," 248–254.

able to simulate genuine friendships, the care-receiver would not experience the full goods of friendship, and it seems that most people would prefer genuine friendship or care over the appearance of them. While the appearance of a friendship may be better than a care-receiver not having care at all, it is evident that a genuine friendship as with a human caretaker is preferable over a care bot's simulation of one. Considering this issue through an Aristotelian lens is beneficial in understanding the significance of why genuine caring is important.

Care bot types such as PARO can be guilty of deceiving the care-receiver because by simulating behaviors such as reacting when being pet and making affective noises, PARO is intentionally simulating a caring relationship by invoking feelings of care in the elders. Since PARO is programmed to act in this manner, the deception is quite literally coded into its function. However, without the deception, it would fail to fulfill its function because it could not provide companionship. Thus, I argue that care bots like PARO are inherently deceptive and thus are morally unfavorable.

The case is quite different for nurse bots, however, whose encoded function is primarily to aid care-receivers with tasks and reminders, and companionship or emotional support is a positive benefit served merely by their physical presence. The distinguishing feature between nurse bots and care bots lies in the fact that nurse bots function to assist the care receivers in practical and physical situations, whereas care bots such as PARO serve to assist the care receivers in emotional situations. As such, nurse bots are not guilty of deception. There are two major ways to circumvent the claim from opponents of this view that nurse bots engage in deceptive practices. The first, which is rather straightforward, draws on the Aristotelian argument that states if one mistakenly believes that a nurse bot is providing genuine care and friendship when it never has, they have not been deceived, but they simply misunderstood the function of the robot in the first place.[89] Thus, nurse bots can be protected from this claim by arguing that their purpose was never to invoke feelings of care in the receiver, but rather simply to assist them.

Further support for the protection of nurse bots against deception is that they, much like actual nurses, simulate behaviors of benevolence, rather than care, and as such, their accompanying companionship cannot be considered deceptive. (T) This argument is supported by drawing a parallel between nurse bots and healthcare professionals (HCPs) such as doctors or nurses. One prominent point of conversation within the medical field regarding HCPs is how invested they should be in the care and lives of their patients. Generally, when considering the high-stress job of HCPs, it is often encouraged for them to not get emotionally involved with their patients for the sake of their own mental health as well as to preserve the objectivity that should be applied to patient treatments. However, when applying the definition of care to HCPs where care calls for the protection of someone and providing what they need, there seems to be an inevitable emotional connotation associated with it. For this reason, prominent philosopher HJ Curzer argues that HCPs should not care for their patients, but rather should have benevolence towards them, meaning they should have a positive emotional attachment to their patients, but

[89] Ibid.

that attachment should be much less than the emotional attachment associated with caring.[90] His argument's support is rooted in prioritizing what is best for the patient, reducing burn-out in HCPs, and increasing equal and fair outcomes. Deep emotional attachment to a patient as seen when genuine care is involved, negatively impacts the patient care delivered because doctors may be more wary to prescribe an appropriate medical treatment that will cause pain or assume a paternalistic attitude towards the patient when they attempt to make decisions for the patient out of their emotional investment.[91] Furthermore, having this emotional attachment has the potential to cause health inequalities because not all patients would receive the same treatment, and perhaps in extreme situations, certain patients would be favored for better treatments. Overall, having an emotional attachment to persons in general over individuals, as seen with benevolence, is essential to providing proper medical care and preserving the emotional health of HCPs, thus, not providing care but benevolence is morally acceptable and ideal. Applying this reasoning to nurse bots, which perhaps are their own form of HCPs, acting with benevolence, not care, is also ideal to ensure the receiver receives the best, unbiased care, ensuring healthcare opportunities stay equal and in the best interest of the patients. Within this perspective, the actions of the nurse bot are not deceptive but rather fall in line with what it means to be benevolent. For these two reasons, care bots such as nurse bots are able to circumnavigate the claim of deception well, demonstrating how their use is ethically acceptable and practically very beneficial.

[90] Curzer, "Is Care a Virtue for Health Care Professionals?," 51–69.
[91] Ibid.

4. Looking to the Future

In conclusion, this paper has demonstrated that while care bots have the ability to revolutionize elder care, there are also accompanying ethical considerations that must be taken into consideration, especially regarding the issues of social isolation, paternalism, and deception. I have attempted to show that certain care bot technologies, such as nurse bots, are morally permissible as the claim of deception does not apply to them; however, deception is quite obvious in other care bot technologies like that of PARO. The importance of parsing out these ethical issues, as I have done above, with deception is essential before care bots can be implemented as a household staple in response to the booming aging population. Care bots surely have a place in the future of healthcare; however, significant ethical work must be done first in order to ensure that human dignity and principles are upheld.

Works Cited

Bureau, U. C. (n.d.). *U.S. Older Population Grew From 2010 to 2020 at Fastest Rate Since 1880 to 1890*. Census.Gov. Retrieved December 4, 2023, from https://www.census.gov/library/stories/2023/05/2020-census-united-states-older-population-grew.html

Care. (2023, December 6). https://dictionary.cambridge.org/us/dictionary/english/care

Coin, A., & Dubljević, V. (2021). "Carebots for eldercare: Technology, ethics, and implications." In *Trust in Human-Robot Interaction* (pp. 553–569). Elsevier. https://doi.org/10.1016/B978-0-12-819472-0.00024-1

Curzer, H. J. (1993). "Is Care a Virtue for Health Care Professionals?" *Journal of Medicine and Philosophy*, *18*(1), 51–69. https://doi.org/10.1093/jmp/18.1.51

Deceive. (2023, December 6). https://dictionary.cambridge.org/us/dictionary/english/deceive

Elder, A. (2016). "False friends and false coinage: A tool for navigating the ethics of sociable robots." *ACM SIGCAS Computers and Society*, *45*(3), 248–254. https://doi.org/10.1145/2874239.2874274

Fernández-Ballesteros, R., Sánchez-Izquierdo, M., Olmos, R., Huici, C., Ribera Casado, J. M., & Cruz Jentoft, A. (2019). "Paternalism vs. Autonomy: Are They Alternative Types of Formal Care?" *Frontiers in Psychology*, *10*, 1460. https://doi.org/10.3389/fpsyg.2019.01460

Knappe, S. (2021). “Dignity or degradation: The risks and realities of carebots in Quebec.” *2021 IEEE International Symposium on Technology and Society (ISTAS)*, 1–8. https://doi.org/10.1109/ISTAS52410.2021.9629175

Koistinen, A.-K. (2016). “The (care) robot in science fiction: A monster or a tool for the future?” *Confero: Essays on Education, Philosophy and Politics*, *4*(2), Article 2. https://doi.org/10.3384/confero.2001-4562.161212

Sharkey, A., & Sharkey, N. (2012). “Granny and the robots: Ethical issues in robot care for the elderly.” *Ethics and Information Technology*, *14*(1), 27–40. https://doi.org/10.1007/s10676-010-9234-6

Sharkey, N., & Sharkey, A. (2012). “The Eldercare Factory.” *Gerontology*, *58*(3), 282–288. https://doi.org/10.1159/000329483

Wilson, R. S., Krueger, K. R., Arnold, S. E., Schneider, J. A., Kelly, J. F., Barnes, L. L., Tang, Y., & Bennett, D. A. (2007). “Loneliness and risk of Alzheimer's disease.” *Archives of General Psychiatry*, *64*(2), 234–240. https://doi.org/10.1001/archpsyc.64.2.234

Yew, G. C. K. (2021). “Trust in and Ethical Design of Carebots: The Case for Ethics of Care.” *International Journal of Social Robotics*, *13*(4), 629–645. https://doi.org/10.1007/s12369-020-00653-w

Humor Against Theodicy

Tristan Latour

The problem of Evil in the face of an omnibenevolent God is simple: how can such an almighty being allow for suffering and injustice? In the past millennia, many thinkers tried to solve that issue: Building a theodicy, a defense of God's perfection, they aimed at exonerating the Supreme Being from causing evil. To counter these attempts, this paper offers a new argument, "from humor," which disproves the perfection of God, and therefore, undermines any foundation for belief in such an entity. Its sole requirement is the very existence of a joke, a laughter, or even a pun. Using the Incongruity Theory of humor, Wittgenstein's aesthetics, the Ireneaen theodicy, and even The Name of the Rose, this paper thus presents an original and definitive objection to any defense of God's perfection in the light of evil in the world. The argument depends on two premises: the perfection of any world created by a perfect God, and the assertion that humor arises from subverted expectations. With these premises in mind, I demonstrate that humor, by showing the failure of our suppositions, reveals a world that often does not fit our needs, does not match our hopes, does not fit human purposes, and thus, fails to earn the designation of "perfect." In a perfect world, humor would be impossible, for all expectations would be correct; no imperfection, no incoherence, no failure would give rise to our humor, because none of these phenomena would exist! Humor reveals an abyss, separating our human conjectures from reality's punchlines. This abyss is an imperfection, often unfit for humanity's needs; and the imperfect God creating a world with such imperfection is unworthy of a capital letter. The argument, which might end fruitless theological battles, will at least bring around the joyful company of our most philosophical ally…humor.

1. Introduction

One of the most common arguments against a theistic worldview is often called the problem of evil (which I will later simplify as the problem of imperfection): How can a perfect and benevolent individual create an imperfect world where evil exists? Theologians of all eras have tried to answer this question by creating a theodicy (a term invented by Gottfried Leibniz, literally a "vindication of God"), an explanation justifying the existence of evil while exonerating God from the blame. To examine these theodicies is crucial to decide whether a perfect God exists or not, whether his nature is good or not, and whether our lives' purpose relies on him or not.

Starting with a discussion of the theodicy proposed by British philosopher of religion John Hick, my aim is to provide a small, yet original help in undermining the belief in a perfect being, that is, in theism. This contribution takes the form of an "Argument from Humor", an anti-theodicy argument of my own invention. I shall now explain the reasoning, the relevance, and the philosophical implications of the argument in question.

2. What is a Theodicy?

Regarding the imperfection of the world, thinkers such as Augustine, in the City of God[92], have initially attempted to deny its actuality by considering evil as a degeneration of God's perfect world. However, this Augustinian approach is strongly problematic, since it challenges the omniscience and omnipotence of God: an all-knowing and all-powerful entity would have foreseen, and been able to prevent, this degeneration into evil. There are many alternative

[92] Augustine, *The City of God*.

theodicies, made by various intellectuals (Leibniz, Origen, Ibn Sina…), but one of the most serious of these solutions, called the Irenaean theodicy, claims instead that God did create evil, purposefully. To analyze this alternative sort of theodicy, I shall bring up Christian philosopher John Hick, and the version he believed in.

For Hick, in accordance with the Bible, evil is real, and can be divided into two categories: moral evil, that God seems to allow, and non-moral evil, that occurs because of the world created by God.[93] To Hick, moral evil is a consequence of the divine gift of free will, so that humans would be able to make a moral choice between good and evil. Hick considers a free decision of that kind, by definition, as causally unexplainable: “The origin of moral evil lies forever concealed within the mystery of human freedom.”[94] On this picture, the justification for moral evil is that God, to test humans, must offer two actual alternatives (good and evil), with parallel consequences (heaven & hell, blissful rest & tragic scourges, etc.), for humans to freely choose. Moral evil allows for freedom, and thus, for deliberately good actions.

Concerning non-moral evil, Hick tries to prove that all the natural disasters, coincidences and accidents which constitute this type of evil are, in fact, serving the purpose of the universe, which he designates as “soul-making.”[95] He argues that these difficult conditions give us the best opportunities to become good, and therefore, to become worthy of God’s love and rewards. Supposedly, a world with a different amount of non-moral evils would thus make our virtues

[93] Found in Pojman, Louis P., et al., “There Is a Reason Why God Allows Evil,” 130.
[94] Ibid, 131.
[95] Ibid, 132-3.

useless: Why need courage if there is no danger? Or why need generosity if no one needs anything? At this point, Hick joins Leibniz's quote: "our world is the best possible world,"[96] if we take it to be the best for "soul-making." Therefore, Hick considers the existence of evil in general, even non- moral evil, as a necessary condition for human morality.

3. The Problem of Imperfection

Many objections have been raised against theodicies, reaffirming the relevance of the problem of evil. Most of them tackle the problem in its narrowest sense, affirming the unnecessary nature of some harmful events, in order to show that God is not excused from such evil. But these objections are problematic: They still follow theistic (usually, Christian) assumptions about "good" and "evil," about moral actions and moral responsibility. The issue is that a critic of theism, such as Spinoza, would in fact dismiss the argument from evil, because it claims the very existence of an inherent, metaphysical "evil" that he did not consider real.[97] The morally Christian framework of the problem of evil, thus mired in trivial moral considerations, is leading the debate astray from the metaphysical discussion of God. This is why the so-called "problem of evil" deserves to be expanded into a "problem of imperfection" in general, which bears the advantage of showing how any imperfect thing, regardless of its morality, is an objection to the existence of God.

Indeed, for every theodicy, it is logically argued that an almighty, all-benevolent and all-

[96] Leibniz, *Theodicy: Essays on the Goodness of God, the Freedom of Man, and the Origin of Evil*.
[97] De Spinoza, *Ethics*. See Preface of Part 4 for his opinion on good and evil.

knowing being would create a perfect world. Even in Hick's Irenaean view, evil remains created by God and therefore, a perfect part of God's perfect plan. God could neither have done otherwise, nor better. If the world is a perfect mechanism, then all its complex parts (even evil) are in their respective, perfect places. Meanwhile, an imperfect god (with an immense, yet finite amount of power, knowledge, or benevolence), such as the members of the pagan pantheons, would be unworthy of our trust, being either unable or unwilling to truly help us.

With the latter assertion, it becomes important to examine the perfection of God in greater detail. God is often defined through perfection, and Anselm's famous ontological proof, for instance, is grounded in such a definition. However, some believers will argue that God does not have to be perfect, that a most powerful, yet imperfect entity would suffice instead. If it were true, then the argument from humor would only disprove the existence of a perfect God, not the veracity of theism itself. However, this line of thought creates an unsuspected, bigger problem for the worshiper of an imperfect God.

Let us suppose that God is, indeed, the most powerful being, despite not having unlimited power (i.e., not being perfect). He therefore has a certain degree of power. But in that case, there could theoretically be an entity (possibly undiscovered yet) reaching a degree of power that is just above God's. Such an entity would thus be more powerful than God; but a God that is not the most powerful entity becomes unworthy of the divine title, being nothing more than "a very powerful entity," and not God. On the other hand, this new, more powerful entity would now deserve the title of God, yet would then fall victim to the very same paradox, ad infinitum…

Thus, the only way for any entity to be considered as God (i.e., as the most powerful entity) is to be infinitely powerful. And since the same problem applies to all of God's usual attributes (power, but also benevolence, and knowledge), then the only possible solution for monotheistic believers is to commit themselves to the perfection of God.

In the face of such reasoning, believers are bound to believe in God's perfection; and this perfect maker, both infinitely capable and infinitely good, would always make the most perfect choices in the creation of his world. If God exists, then the world must be perfect; and here comes the problem of imperfection.

Before we go further, we must also keep in mind that a perfect God implies a perfect world: Anyone claiming that God could have unwillingly created an imperfect world would find themselves denying God's perfect power[98]; on the other hand, anyone claiming that God could have willingly created an imperfect world would simply be denying God's perfect benevolence.[99] Thus, it is inevitable to realize that an imperfect world is incompatible with an almighty, all-knowing, and benevolent (i.e., theistic) God, and that imperfection would indeed disprove the God hypothesis.

[98] An example of such a defense is Augustine's theodicy, where God simply could not prevent the birth of Evil. It appears that even a conception of God as incapable of logical impossibilities (but still omnipotent within the realm of logical possibilities) will not be sufficient to save that theodicy, since nothing in the existence of the Good logically implies the existence of Evil. It would not be a logical contradiction to have Good without Evil, just like there could be light without shadows (if there was light everywhere, for instance). Therefore, any sort of almighty God, logic- bound or not, would have the power to counter the existence of involuntary.

[99] The Theodicy designed by Malebranche falls under that category, arguing that God could have created a perfect world but voluntarily did not do so, to preserve simplicity. A simple reply (along the lines of the above counterargument) would be to point at the contradiction between God's omnibenevolence and this strange concern for simplicity...

4. Humor and Perfection

Yet to do so, one needs to show the existence of imperfections. This is where the existence of humor might constitute a refreshing, decisive argument against theism.

Barely noticeable in the western philosophical tradition, the analysis of humor was often confined to insignificant footnotes in the pages of the classics, until Bergson put the topic forward in his 1900 book on laughter.[100] Since then, philosophers and psychologists alike started wondering about the nature of humor, its place in the human psyche or even its ontological implications; Bergson, for instance, found a way to connect his metaphysical dualism with the mechanisms of humor. At the same time, specialists of all horizons started looking back at the works of older philosophers, tracing the discussion of humor back to Plato and his disdain for laughter (Republic, 388e). Several philosophical theories on the functioning of humor have been held by various thinkers from the past; yet among them, according to the Stanford Encyclopedia of Philosophy,only the Incongruity Theory remains "the dominant theory of humor in philosophy and psychology."[101]

This theory, which states that humor arises from the perception of "something incongruous — something that violates our mental patterns and expectations"[102] — will be the basis for my argument "from humor." There are incongruous situations that lead to more dramatic feelings, depending on personal sensibility and context, yet humor does seem to arise,

[100] Bergson, *Laughter*.
[101] Morreall, "Philosophy of Humor."
[102] Ibid.

in every occurrence where it is felt, from subverted expectations. Therefore, this theory will serve as a basic framework to understand the process of humor: On the one hand, there is an expectation, rooted in psychological assumptions, that tries to predict the outcome of a situation; on the other hand, there is an actual, different outcome, a punchline that breaks this expectation and causes laughter. Humor thus lies in the failure of the perceiver to correctly predict the outcome.

But as we delve further into how humor proves anything about the world, a working definition of perfection needs to be drawn from Wittgenstein's Lectures on aesthetics.[103]

Although he never explicitly speaks of "perfection," the Austrian thinker finds himself explaining how aesthetic judgments come to be made. When talking about musical criticism, he tells his students that "The words [a critic uses] are more akin to 'right' and 'correct' (as these words are used in ordinary speech) than to 'beautiful' and 'lovely.'"[104] He later gives a concrete example: "What does a person who knows a good suit say when trying on a suit at the tailor's? 'That's the right length,' 'That's too short,' 'That's too narrow.'"[105]

Now, is there anything in the notion of "perfection" that is not completely summed up with this illustration? The move from aesthetic appreciation to the judgment of perfection is smooth: What is the "perfect" meal, if not the one correctly fitting the extent of our needs? In archery, what is the "perfect shot" if not the one reaching the center of its target? Perfection is all

[103] Wittgenstein, *Lectures and Conversations on Aesthetics, Psychology and Religious Belief.*

[104] Ibid, 1.8.

[105] Ibid, 1.13.

about the “right” amount, the “correct” length, something Wittgenstein considered our real aim in aesthetics. Perfection is thus relative to a purpose, to an ensemble of criteria, to a certain perspective: What might be a perfect movie, to me, might not fit the precise needs of another member of the audience. Brought back to metaphysics, the notion of perfection is thus applicable to anything that is justly and rightly fitting its expected purpose.

With this definition in mind, it is time to realize that humor, by showing the failure of our expectations, reveals a world that often does not fit our needs, does not match our hopes, does not fit human purposes, and thus fails to earn the designation of “perfect.” In a perfect world, humor would be thoroughly impossible, for all our expectations would be correct; no imperfection, no incoherence, no failure would give rise to our humor, since none of these phenomena could be observed. Humor reveals a gap, separating our human conjectures from reality’s punchline. This gap is an imperfection, often unfit for humanity’s needs; and a world containing such an imperfection is unavoidably imperfect.

One predictable counterargument against this thesis would be a denial of the human perspective about perfection: What if the world was objectively perfect, fitting God’s purpose, and only called imperfect through the lens of our human biases? Two responses can be offered to such an objection. First, it merely transfers the imperfection to “our human biases,” which would still be a flaw in the world, and thus, an imperfection. Then, it remains the case that any flawed perspective is an imperfect thing within the world, fitting neither human purposes, nor God’s omnibenevolent designs, whatever they may be.

Therefore, it becomes clear that humor, showing the weaknesses of our expectations, proves the imperfect nature of the world. To be effective, the smallest pun, the slightest joke needs something wrong or out-of-touch in our processes of cognition; and given our propensity to humor, it appears that the perfection of the world has simply been disproved.

5. The Argument from Humor

With the imperfection of the world now assured, it becomes possible — perhaps, obligatory — to use that knowledge to disprove the existence of God. Thus, the complete argument from humor proceeds as follows:

1. A theistic God cannot have created an imperfect world.
2. Humor shows that the world is imperfect.
3. Therefore, the world cannot have been created by a theistic God.

Premise (1) originates from the incompatibility between an imperfect world and the theistic God, who must be perfect for him to be God at all.

Premise (2), as explained above, is the empirical turning point of the argument. It is based on the inherent imperfection of our understanding of the world, revealed by our mistaken expectations. By extension, this imperfection strips the whole universe from a global, absolute perfection, and leaves that notion to Wittgenstein's realm of relative aesthetics, in which perfection is merely the "right" measure for a certain context. Perfection can apply relatively to certain objects or situations, but not to the entire world, which contains our imperfect cognition. The claims of worldly perfection implied by theodicies (Hick's, Augustine's, Leibniz's) are thus undermined by the presence of imperfection. And furthermore, this devastating problem of imperfection, revealed by humor, has the benefit of surpassing the problem of Evil through its avoidance of inessential moral considerations.

Therefore, from both premises derives the conclusion (3) that the world, imperfect as it is, cannot have been created by the perfect, theistic God. Indeed, if at least one thing is imperfect (which humor demonstrates), then it implies that imperfection does exist. And since no imperfect thing could ever be created by a perfect God, we can assume that this maker cannot be perfect; and an imperfect God does not deserve a capital letter, let alone our faith.

Thus, the undeniable existence of humor, jokes, puns, laughs, and irony is a constant argument against all theodicies, which are doomed by their implied assertion of God's perfection. And if theodicies are all wrong, then it is safe to assume that there is no benevolent demiurge in our interest to worship.

6. Humor as a Tool

With humor directly undermining theism, Abrahamic religions have often struggled against comedy, considering it as a dangerous weapon of evil itself.[106] The devil remains, after all, the one who ridicules God and his plans. The etymology of the word, "devil," happens to mean "the one who divides, who slanders," just as humor reveals the distinction between what reality is, and what our expectations make us project. It is often implied that the works of God deserve

[106] For literary meditations on the theological significance of laughter, one can only recommend Eco, *The Name of the Rose*.

seriousness and solemnity, while critics and laughs are categorized as blasphemies.[107]

The value of humor as a means to reveal any truth about the world might be doubted by the partisans of seriousness. But as a matter of fact, it has always been one of mankind's best ways to understand the actual nature of things. Socrates himself used irony to bring out a constructed, dialectical opinion. The humor of Diogenes the Cynic made his doctrine as remarkable and memorable as Plato's Academy. Through comedy, playwrights such as Aristophanes were able to assert their moral views, just as, two millennia later, Friedrich Nietzsche preferred wordplays and amusing aphorisms because he knew humor to be a valuable means to share philosophical findings. Furthermore, if humor also encompasses the absurd, then Camus's Absurdism is the recognition that value in life cannot be found anywhere else than in the humorous acceptance of meaningless imperfection.

In fact, if we assume the emotional, pathological value of humor (in the Aristotelian sense of pathos), we unveil the reasons why it is such a good guide: It pleases us, by initiating a positive, healthy reaction from the organism; it makes us think, by broadening our intellectual horizons; it allows us to encounter the unsettling chaos of the world while putting it at an emotional distance; it sharpens our critical sense by showing the weaknesses of everything around us. Humor, with all these virtues, appears to be a legitimate philosophical tool, casting a

[107] The few cases where humor (such as Jewish humor) is tolerated by the religious authorities are only made possible by cultural reasons (reaction against oppression, strong reasons to believe in the world's imperfection), and lead either to incoherent, compartmentalized beliefs in both God and cultural humor, or to non-theism (where one abandons the belief in God to fully accept the humor of the culture). The Coen brothers' *A Serious Man* (2009), for instance, offers a tender (but very lucid) account of the difficulties that arise from the former solution.

fresh and elevated perspective upon the walls against which our seriousness stumbles.
Given these insights, if this humorous tool does indicate the impossibility for a theistic account of reality to be true, then we find ourselves choosing between the existence of God and the existence of humor. And since the existence of humor does not need to be proved, then shall we use our old, shiny Occam's Razor and wipe the hypothesis of that God from the picture.

In the end, it appears that theodicies about an almighty, all-benevolent supreme being are defeated by the slightest bit of humor. Maybe is this why the Monty Pythons, in their 1979 masterpiece Life of Brian[108] (widely banned throughout the Christian countries when it came out), chose to end their movie with the actual martyr of religious thinking: humor, personified as the miserable Brian Cohen, who gets crucified by mistake while Jesus himself has previously escaped his execution, thanks to a misunderstanding. Just like humor, nobody intervenes to save Brian, who dies an unfortunate witness of religious mistake. In the end, the only respect we can pay both of them is to cheer up, and laugh…

[108] Jones, Terry, director. Life of Brian. Handmade Films, 1979.

Works Cited

Augustine. *The City of God*. Edited by Demetrius B. Zema et al., Catholic University of America Press, 2008.

Bergson, Henri. *Laughter*. Translated by Cloudesley Brereton and Fred Rothwell, Duke Classics, 2020.

De Spinoza, Benedict. *Ethics*. Edited by Stuart Hampshire, Translated by Edwin Curley, Penguin Classics, 1996.

Eco, Umberto. *The Name of the Rose*. Vintage Classics, 2004.

Leibniz, Gottfried Wilhelm. *Theodicy: Essays on the Goodness of God, the Freedom of Man, and the Origin of Evil*. Edited by Austin Farrer and E. M. Huggard, Anodos Books, 2017.

Morreall, John. "Philosophy of Humor." Stanford Encyclopedia of Philosophy, 20 Aug. 2020, plato.stanford.edu/entries/humor/#IncThe.

Pojman, Louis P., et al. "There Is a Reason Why God Allows Evil." Philosophy: The Quest for Truth, Oxford University Press, New York City, 2020, pp. 129–133.

Wittgenstein, Ludwig, et al. *Lectures and Conversations on Aesthetics, Psychology and Religious Belief: Compiled from Notes Taken by Yorick Smythies, Rush Rhees and James Taylor*. Edited by Cyril Barrett, University of California Press, 2007.

Idealism and Well-Founded Phenomena in Leibniz

Jackson Hawkins

Leibniz maintained that the most real created entities are simple substances called monads, which according to Leibniz are minds or mind-like things. Furthermore, on a common reading of Leibniz, everything in the universe that is not a monad belongs to some inferior level of reality. One of the most important such inferior levels is that of phenomena, which, for Leibniz, are the representational contents of perceptions. This much is uncontroversial. However, an issue in Leibniz's philosophy which has received relatively little direct attention concerns the nature of what he calls "well-founded phenomena." More specifically, very few commentators have discussed what exactly the property of "well-foundedness" might entail. In this paper, I advance a reading of well-foundedness that takes it to be based on what Leibniz calls coherence. In so doing, I argue against an alternative account of well-foundedness that has occasionally been defended by interpreters of Leibniz, according to whom well-foundedness is simply equivalent to the property of representing a real thing.

1. Introduction

Leibniz is commonly understood to have arrived at a type of "phenomenalism" by the end of his philosophical career, of which the best summation may be his assertion that "there is nothing in the world except simple substances, and, in them, perception and appetite."[109] The simple substances referred to in this passage are, of course, Leibniz's monads. In addition to being simple substances, Leibniz considers monads to be minds or mind-like things, as is clear from

[109] Gottfried Wilhelm Leibniz and Leroy Loemker, *Philosophical Papers and Letters,* 2 vols (Chicago: University of Chicago Press, 1956), 1:537.

his remark that "whether these principles of action and of perception are then to be called *forms*, *entelechies, souls* or *minds*… things will not be changed in any way."[110] Leibniz's phenomenalism, as generally understood, thus amounts to the doctrine that the most real created entities are minds or mind-like simple substances that contain perceptions and appetitions and that everything else in the universe belongs to some inferior level of reality. One of the most important inferior levels of reality in Leibniz's system is *phenomena*, a fact which is prefigured in the label "phenomenalism." Indeed, the centrality of phenomena to Leibniz's system is evident from his claim that "in the end, everything reduces to these unities [monads], the rest or the results being nothing but well-founded phenomena."[111]

For Leibniz, phenomena are the representational contents of perceptions, and these contents are deemed metaphysically inferior to substances due to Leibniz's acceptance of the Scholastic maxim that unity and reality are mutually interchangeable properties. As Leibniz puts it, "What is not truly *one* being is not truly one *being* either."[112] For Leibniz, phenomena are unified only in minds and therefore *exist* only in minds. Leibniz's favorite way of illustrating this idea is the rainbow; strictly speaking, a rainbow is a mental representation of a collection of water droplets, which have unity as a single continuous being only when represented as a phenomenon in a mind. And since unity and reality are interchangeable, the rainbow only *exists*

[110] Glenn Hartz, *Leibniz's Final System: Monads, Matter, and Animals* (London: Routledge, 2007), 172.
[111] Gottfried Wilhelm Leibniz, *Philosophical Essays,* trans. Roger Ariew and Daniel Garber (Indianapolis: Hackett, 1989), 147.
[112] Leibniz, *Philosophical Essays,* 86.

in the mind in which it is represented. As Leibniz puts it, "a thing which is aggregated from many things is not one except mentally, and has no reality except that which is borrowed from its constituents."[113]

Although "phenomenalist" readings of Leibniz are common, an issue that has typically been neglected by commentators concerns the qualifier "well-founded" that Leibniz often attaches to the term "phenomenon." In fact, very few interpreters of Leibniz have directly addressed the question of what exactly the property of "well-foundedness" might involve. This reticence may be an effect of the fact that Leibniz himself, though he frequently invokes well-founded phenomena, says comparatively little about the property of well-foundedness per se. Moreover, when Leibniz does address this subject, his comments are frequently somewhat elliptical. The goal of this paper is thus to offer a reading of Leibniz's understanding of well-foundedness, in order to determine what he considers to be fundamental to this property. In so doing, I will argue that the few commentators who have made pronouncements on this issue have tended to ignore the condition that Leibniz himself treats as most important to well-foundedness.

From this point on, I will refer to phenomena that lack the property of well-foundedness as "poorly-founded" for convenience, though this is not Leibniz's preferred term.

2. The Representational Success Reading

[113] Donald Rutherford, "Leibniz's 'Analysis of Multitude and Phenomena into Unities and Reality,'" *Journal of the History of Philosophy* 28, no. 4 (1990): 537.

Although well-foundedness has attracted relatively little attention, some scholars have put forward claims about it. Donald Rutherford, for instance, has argued that well-foundedness is the property of representing a real object. According to Rutherford, this is the only way to make sense of Leibniz's claim that bodies are aggregates of monads.

> An analysis of the content of corporeal phenomena reveals them to be perceptions of other monads… Only in this case, I would argue, is the notion of body as a "well-founded" phenomenon analyzed in such a way as to make sense of Leibniz's abiding commitment to the thesis that bodies are aggregates of monads.[114]

In brief, Rutherford argues that the only intelligible way to understand Leibniz's claims that extended bodies are aggregates of unextended monads is to read him as claiming that the extendedness of bodies is an illusion built out of minds' confused perceptions of collections of unextended monads, in much the same way that a rainbow is a mental interpretation of a collection of water droplets. In this sense alone, Rutherford claims, are bodies "aggregates" of monads. Rutherford further thinks that this reading necessitates the conclusion that the well-foundedness of phenomena just *is* the property of being a representation of real things, namely, monads.

Similarly, Shane Duarte has claimed that "it seems clear that Leibniz understands a well-founded phenomenon to be the representational content of a perception that has an extra-mental object."[115] Duarte, however, arrives at this conclusion because he understands

[114] Donald Rutherford, "Phenomenalism and the Reality of Body in Leibniz's Later Philosophy," *Studia Leibnitiana* 22, no. 1 (1990): 27.

[115] Shane Duarte, "The Ontological Status of Bodies in Leibniz (Part I)," *Studia Leibnitiana* 47, no. 2 (2015): 148.

Leibniz to consistently employ the Scholastic distinction between a thing's extra-mental existence (existence *a parte rei*) and its existence as a mental representation (existence *quoad nos*). According to Duarte, a phenomenon is well-founded when its existence *quoad nos* is grounded in the existence *a parte rei* of some real entity.

From this point on, I will call this the *representational success* reading, since Duarte and Rutherford both maintain that the property of well-foundedness is equivalent to the property of successfully representing a really existing thing. I will argue, however, that Leibniz himself identifies an entirely different condition as fundamental to well-foundedness. One might circumscribe this overlooked condition within the label "coherence." In a later section, I will give a more detailed response to Duarte and Rutherford's respective versions of the representational success reading, but before doing so, I will put forward my own understanding of well-foundedness that, I contend, comports more readily with Leibniz's comments on the topic.

3. Coherence and Metaphysico-Mathematical Agreement

Leibniz's most revealing statement on the issue of well-foundedness may lie in a letter to Giambattista Tolomei:

> Extension, and in it bulk or impenetrability… are in fact, I hold along with many ancient thinkers, only well-founded phenomena: certainly not phenomena that deceive but phenomena that have nothing else objectively real except that by which we distinguish dreams from waking, which is to say, the metaphysico-mathematical agreement among themselves of all those things which souls or entelechies perceive, whether you compare

> these phenomena with themselves in the same entelechy or compare them with the phenomena of other entelechies.[116]

In this passage, Leibniz makes several noteworthy claims:

1. There is nothing to distinguish well-founded phenomena from poorly-founded phenomena except that by which dreams are distinguished from wakeful states.
2. This distinction is made by means of a "metaphysico-mathematical agreement" possessed by well-founded phenomena.
3. The metaphysico-mathematical agreement that distinguishes well-founded from poorly-founded phenomena can be observed both in individual phenomena and through the comparison of multiple phenomena.

Despite its apparent centrality to Leibniz's understanding of well-foundedness, the meaning of the phrase "metaphysico-mathematical agreement" is somewhat opaque. However, valuable insight into what Leibniz might intend here can be gleaned from his much earlier treatise, "On the Method of Distinguishing Real from Imaginary Phenomena" (MRI). Although in this text Leibniz speaks of "real phenomena" and "imaginary phenomena," I will assume that the distinction between real and imaginary phenomena is simply an early version of the distinction between well-founded and poorly-founded phenomena, and that the terms involved are more-or-less synonymous. Commentators on Leibniz have generally been willing to permit this exegetical move in light of the prominent similarities between the language of MRI and that of

[116] Gottfried Wilhelm Leibniz, "Leibniz to Giambattista Tolomei," trans. Donald Rutherford, 2014, https://dss-sites.ucsd.edu/drutherford/Leibniz/translations/TolomeiG.pdf.

Leibniz's later writings on well-founded phenomena. For instance, as I will soon show, Leibniz treats *dreams* as paradigmatic examples of both imaginary phenomena and poorly-founded phenomena.

Importantly in MRI, Leibniz describes a number of criteria by which phenomena can be determined to be well-founded. These criteria encompass both considerations of a phenomenon's internal properties and comparisons of multiple phenomena, echoing Leibniz's claim to Tolomei. Leibniz names three strictly internal criteria: vivacity, complexity, and coherence. The first two criteria are fairly simple in scope: "[A phenomenon] will be vivid if its qualities… appear intense enough. It will be complex if these qualities are varied and support our undertaking many experiments and new observations."[117] The idea here is that a phenomenon's qualities must be well-defined and varied enough for that phenomenon to be meaningfully investigated; if a phenomenon is too vague, hazy, or simple to lend itself to experimentation, then it is not well-founded. In comparison to this, the criterion of coherence is far more involved. Leibniz writes that a phenomenon will be coherent

> If it conforms to the customary nature of other phenomena which have repeatedly occurred to us, so that its parts have the same position, order, and outcome in relation to the phenomenon which similar phenomena have had. Otherwise phenomena will be suspect, for, if we were to see men moving through the air astride the hippogryphs of Ariostus, it would, I believe, make us uncertain whether we were dreaming or awake.[118]

[117] Leibniz and Loemker, *Philosophical Papers and Letters,* 2:603.

[118] Leibniz and Loemker, *Philosophical Papers and Letters,* 2:603–4.

This is Leibniz's understanding of the *internal coherence* of a phenomenon. That is, a phenomenon is internally coherent if it resembles (in position, order, and outcome) other, similar phenomena that a substance has previously encountered. Thus, the sight of men riding hippogryphs is internally incoherent because it bears no such resemblance to anything the perceiving substance has hitherto experienced. Of course, in a certain sense, this criterion involves a sort of comparison between the phenomenon in question and the entire ensemble of phenomena that a substance has previously encountered. However, the important point here is that, when this criterion is employed, phenomena are evaluated on the basis of the *resemblance* of their strictly internal properties to those of other phenomena. Conversely, Leibniz also thinks that the criterion of coherence can be evaluated on the basis of a phenomenon's *causal* relations to other phenomena. I will call this the criterion of *external coherence*.

> This criterion can be referred back to another general class of tests drawn from preceding phenomena. The present phenomenon must be coherent with these if, namely, it preserves the same consistency or if a reason can be supplied for it from preceding phenomena or if all together are coherent with the same hypothesis.[119]

The criterion of external coherence adverts to a phenomenon's causal continuity with the phenomena preceding it; if a phenomenon appears "out of the blue," with no discernible connection to the phenomenon preceding it, then it is externally incoherent. This criterion also applies in the opposite temporal direction, to the *predictivity* of a phenomenon with respect to future phenomena. In fact, Leibniz suggests that predictivity is the "most powerful" of all criteria

[119] Leibniz and Loemker, *Philosophical Papers and Letters,* 2:604.

hitherto listed: "Yet the most powerful criterion of the reality of phenomena, sufficient even by itself, is success in predicting future phenomena from past and present ones."[120] Leibniz thus gives his reader a useful set of criteria by which well-foundedness can be ascertained:

1) Vivacity: A phenomenon's qualities must be sufficiently intense to be investigated via experimentation.
2) Complexity: A phenomenon must contain sufficient detail to be investigated via experimentation.
3) Internal coherence: A phenomenon must resemble other phenomena that a mind has previously encountered.
4) External coherence: A phenomenon must be causally continuous with past phenomena and predictive of future phenomena.

In addition to these four criteria, a fifth can be surmised from Leibniz's other writings, though it does not appear overtly in MRI. I will call this criterion 5) *inter-subjective coherence*.

> God could give to each substance its own phenomena independent of those others, but in this way he would have made as many worlds without connection, so to speak, as there are substances, almost as when we say that, when we dream, we are in a world apart and that we enter into the common world when we wake up.[121]

Simply put, this fifth criterion requires that the well-founded phenomena of a certain substance be harmonious with the phenomena of every other substance that exists in the same world. If one of the phenomena in a substance, x, were to contradict the phenomena of the other substances in

[120] Leibniz and Loemker, *Philosophical Papers and Letters,* 2:604.

[121] Leibniz and Loemker, *Philosophical Papers and Letters,* 2:802.

x's world then that phenomenon would not be well-founded. For instance, if x dreams that it perceives a green sky, the phenomenal content of this perception would be disharmonious with the contents of the perceptions of all the other substances in x's world, who perceive the sky as blue. Thus, whereas criterion 4 has to do with the causal consistency of the past, present, and future phenomena of a *single* substance, criterion 5 concerns the comparative harmoniousness of the phenomena of *multiple* substances.

These five criteria for well-foundedness help to clarify what Leibniz might have in mind when he speaks of "metaphysico-mathematical agreement." While criteria 1–3 concern the internal properties of individual phenomena, and thus do not pertain directly to any sort of "agreement" between phenomena, criteria 4 and 5 do advert to such agreement; criterion 4 involves the continuity/predictivity of different phenomena within one substance, and criterion 5 involves the inter-subjective harmony of phenomena across multiple substances. Importantly, in certain texts, Leibniz suggests that the agreement emphasized in these latter criteria can be understood as obedience to the rules of mathematics. For instance, he writes, "Although mathematical meditations are ideal, this does not diminish their utility, for actual things do not depart from mathematical rules. Indeed, one can say that in this consists the reality of phenomena, which distinguishes them from dreams."[122] It would thus appear that the

[122] Gottfried Wilhelm Leibniz and Carl Immanuel Gerhardt, *Die Philosophischen Schriften von Gottfried Wilhelm Leibniz,* 7 vols. (Berlin: Weidmann, 1875), 4:569. My translation. The original reads: "*Quoique les méditations Mathématiques soient idéales, cela ne diminue rien de leur utilité, parce que les choses actuelles ne sauraient s'écarter de leurs règles; et on peut dire en effet, que c'est en cela que consiste la réalité des phénomènes, qui les distingue des songes.*"

“metaphysico-mathematical agreement” that Leibniz invoked in his letter to Tolomei refers to those coherence criteria that involve the obedience of *metaphysical* entities (phenomena) to *mathematical* rules. This applies especially to criterion 5, for it seems that, according to Leibniz, obedience to a common set of rules is partially what ensures that the phenomena of various substances will harmonize with one another.

4. Response to Rutherford and Duarte

An important upshot of Leibniz’s reflections in MRI is the fact that *none* of the criteria for well-foundedness given in the text require that a well-founded phenomenon be representationally successful. On the contrary, Leibniz states that “even if this whole life were said to be only a dream, and the visible world only a phantasm, I should call this dream or this phantasm real enough if we were never deceived by it when we make good use of reason.”[123] Of course, Leibniz does not suggest in this passage that the phenomenal world *is in fact* a mere phantasm, but only that he cannot at present be certain that it is not. However, it is at least clear from this remark that Leibniz would in principle be willing to regard a sufficiently orderly phenomenon as well-founded *even if* it were a phantasm that lacked a real object.

Rutherford’s endorsement of the representational success reading of well-foundedness is a consequence of his efforts to decipher Leibniz’s claims that extended bodies are aggregates of unextended monads. Rutherford thinks that the only way Leibniz can be intelligibly understood on this point is by maintaining that bodies are only aggregates of monads in the sense that they

[123] Leibniz and Loemker, *Philosophical Papers and Letters,* 2:604.

are mental representations of collections of monads. This, Rutherford contends, suggests that well-foundedness is the result of phenomena being grounded in the reality of the monads they represent.

And yet, as we have seen, Leibniz affirms that he would call a representationally-failed phenomenon well-founded, provided that it met certain coherence conditions. It may very well be the case, as Rutherford argues, that we need a notion of representational success to make sense of bodies being aggregates of monads, but it is a leap to claim, from this, that we should also understand well-foundedness on the basis of this representational success. For my part, I am willing to acknowledge that representational success is probably *sufficient* to make a phenomenon well-founded, but Leibniz's writings indicate that it is not *necessary*. Rather, representational success is simply a means of ensuring that a phenomenon will satisfy the more fundamental criteria of coherence canvassed in the previous section. This is because, for Leibniz, the world itself (and every part of the world) is perfectly harmonious and internally coherent. It thus seems that if a phenomenon succeeds in representing a part of this harmonious world, it will be (a) internally coherent, since the object of its representation is necessarily internally coherent and (b) externally and inter-subjectively coherent with all other phenomena which represent things in the same world, since there are no contradictions between various parts of the world. Nevertheless, Leibniz is emphatic that representational success is not the *only* way of ensuring that a phenomenon will meet these coherence criteria and that certain representationally-failed phenomena could still qualify as well-founded if they manage to be coherent enough. The question of whether any coherent but representationally-failed phenomena actually exist is, for present purposes, irrelevant.

Of course, Rutherford is well aware of the passages in which Leibniz seemingly endorses a coherence-based reading of well-foundedness. However, Rutherford waves these remarks aside by branding them "ambiguities."

> [Leibniz] maintains both that phenomena are well-founded because they are "in agreement," and that their foundation is a consequence of each perceiver's being a "mirror" of a common universe of monads… I would argue that the only way to make sense of these comments is to accept that Leibniz allows a considerable degree of ambiguity in the meanings of key metaphysical terms.[124]

It seems uncontroversial that, wherever possible, historical readings should avoid resolving difficulties simply by appealing to ambiguity in primary texts. I thus view my own account as a way of avoiding this move by treating representational success as a *sufficient*, but not *necessary* condition for well-foundedness.

As with Rutherford, Duarte views the representational success reading as a consequence of a broader interpretation of Leibniz. Unlike Rutherford, however, Duarte emphasizes the Scholastic distinction between existence *a parte rei* and existence *quoad nos*, and maintains that, for Leibniz, phenomena are well-founded when their existence *quoad nos* is grounded in the existence *a parte rei* of their representational objects. I make much the same response to Duarte's argument as to Rutherford's, namely, that representational success — and, by the same token,

[124] Rutherford, "Phenomenalism and the Reality of Body in Leibniz's Later Philosophy," 24.

groundedness in the existence *a parte rei* of something in the world—is sufficient but not necessary for well-foundedness.

Of additional interest is Duarte's use of MRI as a proof text for his version of the representational success reading.

> As Leibniz makes plain in his "De modo distinguendi phaenomena realia ab imaginariis," he understands a real phenomenon to be the representational content of a perception that has an extra-mental object. Indeed, the principal aim of this work is to identify criteria or signs (*indicia*) by which one can distinguish those perceptions which have extra-mental objects from those perceptions which do not.[125]

I disagree with this interpretation of MRI. The aim of the treatise is indeed to identify a set of indicia by which real phenomena can be distinguished from imaginary phenomena, but nowhere in the text does Leibniz suggest that he understands this to be a question of which phenomena have extra-mental objects and which do not. In fact, Leibniz is fairly straightforward in stating that he considers the indicia identified in MRI to be incapable of demonstrating the existence of extra-mental objects: "By no argument can it be demonstrated absolutely that bodies exist, nor is there anything to prevent certain well-ordered dreams from being the objects of our mind."[126] It thus seems that, far from vindicating the representational success reading, Leibniz's comments in MRI actually call it into question.

5. Conclusion: Idealism and Pre-Established Harmony

[125] Duarte, "The Ontological Status of Bodies in Leibniz (Part I)," 148.

[126] Leibniz and Loemker, *Philosophical Papers and Letters,* 2:604–05.

At this point, I have not argued that well-founded phenomena are never representationally successful; I have merely shown that Leibniz does not treat representational success as necessary for well-foundedness. However, throughout Leibniz's writings there is a powerful pull in the direction of the more radical thesis that well-founded phenomena are never representationally successful, and that *all* phenomena are mere dreams differentiated by degrees of coherence. This tendency is evident from many of Leibniz's later writings, wherein he overtly entertains this possibility. For example:

> If that substantial bond of monads were absent, then all bodies with all their qualities would be only well-founded phenomena, like a rainbow or an image in a mirror—in a word, continuous dreams that agree perfectly with one another; and in this alone would consist the reality of those phenomena.[127]

This radical thesis is also strongly implied by Leibniz's "windowless" doctrine, which he formulates as follows: "Monads have no windows through which something can enter or leave. Accidents cannot be detached, nor can they go about outside of substances… Thus, neither substance nor accident can enter a monad from without."[128] Perception understood (as in Scholastic Aristotelianism) as consisting of the perceived object impinging on the perceiving substance by imparting an intelligible species to it would be precisely what the windowless doctrine is supposed to prohibit. But if a substance's perceptions are not obtained through interaction with the external world, then the phenomena that constitute the representational

[127] Gottfried Wilhelm Leibniz, *The Leibniz-Des Bosses Correspondence,* trans. Brandon Look and Donald Rutherford (New Haven: Yale University Press, 2007), 227.
[128] Leibniz, *Philosophical Essays,* 214.

contents of these perceptions have no real intercourse with an extra-mental world. In this sense, then, no such phenomena would be representationally successful.

This reading of Leibniz as having gone beyond mere phenomenalism and into full-throated idealism has received its most thorough defense in Robert Adams's *Leibniz: Determinist, Theist, Idealist*. Ultimately, even if one resists the idea that Leibniz fully embraces such idealism, it is undeniable that some of his philosophical commitments incline strongly in that direction. Prominent among these is the fact that, as I have shown above, the property of well-foundedness does not necessarily involve representational success. If this conclusion is accepted, then it is at least possible that for Leibniz all phenomena—including those that are well-founded—are representationally-failed dreams, differentiated by degrees of coherence.

Works Cited

Adams, Robert Merrihew. *Leibniz: Determinist, Theist, Idealist.* Oxford: Oxford University Press, 1994.

Duarte, Shane. "The Ontological Status of Bodies in Leibniz (Part I)." *Studia Leibnitiana* 47, no. 2 (2015): 131–61.

Hartz, Glenn. *Leibniz's Final System: Monads, Matter, and Animals.* London: Routledge, 2007.

Leibniz, Gottfried Wilhelm. "Leibniz to Giambattista Tolomei." Translated by Donald Rutherford. 2014. https://dss-sites.ucsd.edu/drutherford/Leibniz/translations/TolomeiG.pdf.

Leibniz, Gottfried Wilhelm, and Carl Immanuel Gerhardt. *Die Philosophischen Schriften von Gottfried Wilhelm Leibniz.* 7 vols. Berlin: Weidmann, 1875.

Leibniz, Gottfried Wilhelm. *The Leibniz-Des Bosses Correspondence.* Translated by Brandon Look and Donald Rutherford. New Haven: Yale University Press, 2007.

Leibniz, Gottfried Wilhelm. *Philosophical Essays.* Translated by Roger Ariew and Daniel Garber. Indianapolis: Hackett, 1989.

Leibniz, Gottfried Wilhelm, and Leroy Loemker. *Philosophical Papers and Letters.* 2 vols. Chicago: University of Chicago Press, 1956.

Rutherford, Donald. "Leibniz's 'Analysis of Multitude and Phenomena into Unities and Reality.'" *Journal of the History of Philosophy* 28, no. 4 (1990): 525–52.

Rutherford, Donald. “Phenomenalism and the Reality of Body in Leibniz’s Later Philosophy.” *Studia Leibnitiana* 22, no. 1 (1990): 11–28.

Law, Liberty, and The Limits of Selfhood

Adam Lewis Sebastian Lehodey

Is the state justified in protecting individuals from themselves? This paper advances philosophical conversations around the interlinked nature of selfhood and the law, proposing that the self ought to be understood not as an isolated concept, but rather as a series of narratives deeply connected to the communities around us. From this conception of selfhood that is advanced, an analysis of the relationship between individuals, government and the community is put forth, culminating in the consideration of questions surrounding 'consent of the governed.' This paper contributes to the literature on selfhood and the scope of the law by putting political philosophers in conversation with one-another and with decisions made in courthouses over the past century. While ultimately arguing that the state does have a right to protect individuals from themselves in certain cases, it provides a more grounded justification for doing so and calls for a re-evaluation of current policy to ensure it adheres to the principles laid forth.

On one side of the intellectual boxing ring is John Stuart Mill, who claims that individuals are 'not accountable to society for [their] actions, in so far as these concern the interests of no person but [themselves].'[129] On the other side of the ring are thinkers like Richard Thaler and Cass Sunstein, arguing that the state should play a more active role in nudging people towards actions deemed beneficial to their overall wellbeing.[130] The idea of a personal sphere free from state interference isn't new: we see it in ancient Roman family structures, helmed by a powerful *Pater*

[129] Mill et al., *On Liberty, Utilitarianism, and Other Essays*, Chapter 5.
[130] Thaler and Sunstein, *Nudge: Improving Decisions About Health, Wealth, and Happiness*.

Familias (father figure) who held absolute power, including the power of life and death, over his family. The concept is most evident in the American Constitution, which demarcates the individual sphere from the collective through its non-exhaustive enumeration of rights. What's interesting is that paternalism isn't a rejection of the individualist values underpinning the constitution. It isn't the state telling individuals to act against their own interests in favor of that of the *collective* interest. Instead, paternalism amounts to an assertion that individuals should be forced to act in a particular way because it is in their *own* interests to do so.

This essay is about selfhood and the state. More precisely, it is interested in the question of whether the proper role of the state extends to protecting individuals from themselves. The issue has direct salience in light of ongoing debates over access to assisted suicide, drug legalisation, bans on fast-food advertising and, in the case of the UK, moves to ban cigarettes even for consenting adults.[131] For these debates to be more than a battle of *wills,* further analysis is needed.

Preceding the political philosophical debates on consent of the governed and individual rights is a metaphysical debate on what selfhood actually *is* and requires. In policy and beyond, I argue, the self is often conceptualized as being this discrete 'authentic' entity, something which directly justifies the legal recognition and enforcement of rights and would suggest the answer to this paper's research question is *no.*

[131] 'Prime Minister to Create "Smokefree Generation" by Ending Cigarette Sales to Those Born on or after 1 January 2009,' GOV.UK.

This paper posits that this conception of selfhood is misplaced: people don't exist in a vacuum and so the 'individual' must be understood in the context of the broader social and political community into which he is born. This still paves the way for the creation of an 'individual' sphere in the eyes of the law, stemming from a need to manage and mitigate conflict. But that does not necessarily rule out protecting individuals from actions deemed harmful to their welfare. Consent of the governed need not apply to every single action, that would simply be unworkable given the millions of collective decisions that need to be made and the plurality of different interests at stake. Instead, where consensus (*ideally*, or general agreement *in practice*) is needed is on the higher-order rules and frameworks that govern political decision-making.

1. Toward a New Conception of the Self

In our everyday lives, we take it as a given that we, as individuals, are self-defined and well-ordered units, distinct from the rest of the world. In the context of law and justice, this is certainly the dominant doctrine. Baltimore v. Goodman (1927), for example, stated the need for *individuals* to take reasonable precautions in negligence cases.[132] If an individual is convicted of murder, it is he, not his mother that will be incarcerated. The principle is omnipresent in our lives: In economics many of us consider what is just to be what has been meritocratically acquired by individuals through their *own hard work.* Meanwhile our philosophy has progressed

[132] "Baltimore & Ohio R. Co. v. Goodman, 275 U.S. 66 (1927)," Justia Law.

to a view of man whereby, in the words of Schopenhauer, his 'mind is by its nature free, not a slave; only what it does by itself and willingly is successful.'[133]

It is from this perspective of an authentic, true self that we derive rights-based theories of justice, including those proposed by Nozick, Mill and Ayn Rand. In her essay entitled *Man's Rights*, Rand argues that because man exists, and if he is to continue existing, he has a *right* to his own life and by derivation, his own body. For the state to step in and coerce an individual to act — even if in their own interest — against their will, amounts to a violation of man's rights. Rand herself never directly addressed the issue of 'paternalism,' the view that the state should urge people toward behaviors that might advance their welfare. Nevertheless, the work of her intellectual heirs provides some confirmation of this view. One former Fellow of the Ayn Rand Institute went as far as to label anti-smoking legislation a 'cancer on American Liberty' in a 2010 op-ed.[134]

More broadly, underlying this view of the well-defined individual is the concept of a *will* that is perfectly rational and knows exactly what it wants. From this, it follows that when government claims to act in an individual's interest, it is really just infringing upon their rights, using their 'well-being' as an excuse. After all, an individual with perfect rationality and clear desires has no need for government to act on its behalf. Most libertarians, including Rand and Nozick, would concede that governments have the ability to act when a genuine collective action

[133] Schopenhauer, *The World as Will and Representation*.

[134] 'Anti Smoking Paternalism A Cancer on American Liberty,' The Ayn Rand Institute.

problem is involved. However, this could not be extended to cases where no externality is present (or is minimally present), as in the case of someone drinking alone at home. The main takeaway is that this conception of selfhood leaves no space (and no need) for the state to protect individuals from themselves.

The obvious objection is that the individual *can't* be as clearly defined as the view above would imply. Alasdair MacIntyre is one figure who provides a counter-narrative of the self in his 1981 book *After Virtue*.[135] We do not exist in a sandbox but rather as embodied members of a community which shapes our values and vice-versa, he posits. MacIntyre's work draws heavily on that of Aristotle's. The line between the self and his community is far more blurred than it appears, for first is the question of values, which derive from one's community and whom one in part shapes. Then there is a question of ethics: What is considered ethical by a community, even if one does not agree, will shape one's incentives to act in a certain way (take, for example, the age of consent which varies across regions and time periods, but which carries severe penalties for breaching it). From this view, we see things completely differently — things are less a question of the state protecting individuals from themselves, but rather the community taking steps to safeguard its own existence and moral integrity. Actions matter and influence others. There are no neutral acts — everything sends a normative signal. If any individual, under this view, wishes to do good, he must aim at the good of his community. This view seems convincing but falls short. Ultimately, decisions are still being made from an individual even if influenced by

[135] MacIntyre, *After Virtue: A Study in Moral Theory*, Chapter 14.

others. Attempting to drive policy purely based on the majority will therefore *inevitably* lead to conflict and stagnation.

There's another objection we must take seriously: that there is no such thing as the self at all. If we extend the logic of the MacIntyre / Aristotle argument above, we realize that everything we are, both our physical bodies as well as our souls, ideas and minds, are deeply interwoven with the world around us. '*For dust thou art, and to dust thou shalt return*,' as the verse from Genesis goes. This monist understanding deserves credit, but in the words of Parmenides, we life according to 'the way of mortals'[136] and therefore hold onto a pluralistic ideal of the self. The self exists, I have written elsewhere, not as an objective or atomised entity, but rather as a set of narratives one creates about one's life and one's identity.[137]

It has become clear that any protection of individuals against *themselves* cannot proceed on a pure rights theory. We have seen that the self is in fact a fluid concept, shifting over time and being deeply connected to the community. Only from this understanding of the self can we establish the proper limits of government.

2. From Selfhood to Nationhood

Our conversation proceeds from this new paradigm for the self that we have established: One where far from being an atomised unit, 'selfhood' is understood as constructed through narrative

[136] Parmenides, Fragments.

[137] Lehodey, "Decoding the Self through Auster's City of Glass | The New York Trilogy Analysis."

yet still containing an element of autonomy. Individuals do have wills, even if that *will* is not absolute.

To proceed from here, a more thorough investigation of the self and the polis is needed. Individuals, we have recognised, do not live in isolation; and the very morals and standards that individuals assume in their lives are shaped by those of the community. Assuming that individuals are self-interested and rational,[138] it nonetheless holds that attempts to improve themselves will include those aimed at improving society. One cannot live well without those around themselves living well. This reordering in how we understand the self — a view of the individual closely aligned with Aristotle's[139] — does imply that to pursue goodness, individuals must order others around them towards the *good*. Of course, everyone has a different conception of what they take *the good* to mean, and so the result *is* in fact a relativistic majoritarian imposition on other people. This is why, in the case of drugs and alcohol, some countries fix the drinking age at 18 whilst others fix it at 21. All the while Oregon decriminalizes all drugs whilst the UK clamps down.[140]

If this view of justice as a social dynamic seems familiar, it's because it is – this was the perspective of law that dominated before the Enlightenment, visible in Miller's *Crucible* where the villagers of Salem burn witches for the external moral corruption they inflict.[141]

[138] Dawkins, *The Selfish Gene: 40th Anniversary Edition*.
[139] Aristotle, et al. *The Nicomachean Ethics*.
[140] 'Possession of Nitrous Oxide Is Now Illegal,' GOV.UK.
[141] Miller, *The Crucible: A Play in Four Acts*.

As the Enlightenment caught on, so did notions of responsibility and ideas about universal human rights. Flourishing in our own lives, we realized, requires the codification of rights into the law. At the time the dominant rhetoric was commonly one of rights endowed by a creator. The Declaration of Independence, for example, famously states that ‘‘we hold these truths to be self-evident, that all men are created equal, endowed by their Creator with certain unalienable rights.”

First came Rousseau, who argues that individuals enter into social contracts because it is beneficial to *them* to do so, and that Governments are only valid insofar as there is a covenant between men.[142] Although, in *The Social Contract,* Rousseau provides us with a useful perspective on why we should accept government in our lives, the book provides no answer to the limits of said government (indeed he argues that if men choose not to accept, they should be *forced* to ‘be free’). Here, Hayek comes to our rescue, arguing that protecting minority rights is in the interests of all, including the majority, for it is from there which progress is derived, and a progressive society is fundamental to living a good life.[143] Hayek’s claims are complemented by those of Amartya Sen, who illustrates that the value of rights is not purely procedural, but also grounded in their outcomes.[144] We can therefore understand rights as procedures that help to secure the best outcome for the most individuals across an extended period of time, counteracting the Randian and Nozickian argument that rights *exist* out there.

[142] Rousseau, *The Social Contract*.
[143] Hayek and Stelzer, *The Constitution of Liberty.*
[144] Sen, *Development as Freedom*.

Chief amongst these rights which guarantee human flourishing is the right to govern one's life, which implies the individual be left alone so long as they are not harming others. Though one could argue that *no* actions are purely individualistic given the nature of individuals identified further up in this essay, Hayek again notes the need for a useful threshold before which the state can intervene — historically when individuals begin to cause physical harm to others. We find ourselves back at Locke's initial argument for autonomy and consent of the governed, albeit with a much richer understanding of the self and its relation to other selves and the world. To rule without the consent of the governed is to pave the way for despotism and conflict. Only in accordance with this almighty principle can we achieve a state of flourishing in our lives and in those of others.

3. Neutrality and Consent

All of the above points in one direction: Government cannot have either the duty or the right to protect individuals from themselves for this would violate the principle of consent of the governed. I will reiterate that whilst any government *could* of course choose to violate this principle, we are assuming that individuals are self-interested and rational, which therefore limits this possibility.

The principle of the consent of the governed would be violated by asserting that an action protects an individual from themselves. To justify such an action, the individual must recognize the need for protection, thus placing them in the best position to make the decision independently. Alternatively, even if both the individual and government officials agree on intervention, if the government acts on behalf of the individual, it may coerce another party to dedicate time or resources involuntarily. This would occur because, in the absence of state intervention, the transaction would have been purely voluntary. The conclusion is straightforward: the state should neither protect individuals from themselves nor force third parties to contribute to such protection against their will.

But what if there were a way to bridge the two? Peter de Marneffe's paper, *Liberalism, Liberty, and Neutrality* does exactly that. In distinguishing between 'Concrete Neutrality' and 'Neutrality on Grounds,' De Marneffe helps to show that consent can be secured even if individuals do not agree with the outcomes of justice.[145] In a system of law for example, an individual convicted of a crime might not be content with that decision, even if he would concede that the legal system at large is premised on principles with which he agrees. The same is true in a wider system of Government — one need not agree with every single law, but so long as individuals agree with the principles according to which laws are made and justice applied, there is legitimacy in the system. Testing the criteria for if neutrality of grounds has been met is difficult in practice, and we revert to proxy measures like voter turnout and media engagement. Nonetheless, de Marneffe's paper is crucial in advancing our understanding of this question and helping us recognise that the principle of consent can still be met even when individuals do not agree with every specific law.

[145] De Marneffe, "Liberalism, Liberty, and Neutrality," 253–74.

4. Conclusion

We end where we began, at our thesis, having shown that despite selfhood being in many ways an illusion, consent of the governed is still an essential characteristic of any government. From there, I outlined that consent of the governed does not *on principle* exclude the state from protecting individuals against themselves. The exception to this would be if individuals enacted a constitution that outlawed this, or clearly showed their disavowal of these measures in the press or at the ballot-box. Such would be a clear example of the fact that individuals did reject such measures at a second-order level. Failing that, state action aimed at protecting individuals from themselves, such as prohibiting drugs, mandating seatbelts, or outlawing underage drinking, must be assessed on grounds of expediency and not principle.

Having drawn on many thinkers and objects of analysis throughout this essay, I too have advanced the conversation further by providing a stronger justification for rights through the synthesis of various thinkers, causing us to question this issue deeply. I will conclude by urging my readers to think about what good policy on grounds of expediency *means*. It increasingly looks as if the prohibition approach, particularly on drugs and other substances, has failed to deliver over the past few decades. Perhaps now is the time to assert a new path forward.

Works Cited

Aristotle, et al. *The Nicomachean Ethics*. Oxford University Press, 2009.

Aristotle, et al. *The Politics*. Oxford University Press, 2009.

“Baltimore & Ohio R. Co. v. Goodman, 275 U.S. 66 (1927).” Justia Law, https://supreme. justia.com/cases/federal/us/275/66/

Dawkins, Richard. *The Selfish Gene*: 40th Anniversary Edition. Oxford University Press, 2016.

Genesis 3:19. King James Version.

“Jacobson v. Massachusetts, 197 U.S. 11 (1905).” Justia Law, https://supreme.justia.com/cases /federal/us/197/11/.

Jr, Cleve R. Wootson, and Jaclyn Peiser. “Oregon Decriminalizes Possession of Hard Drugs, as Four Other States Legalize Recreational Marijuana.” Washington Post, 4 Nov. 2020. www.washingtonpost.com, https://www.washingtonpost.com/nation/2020/11/04 /election-drugs- oregon-new-jersey/.

Lehodey, Adam Louis Sebastian. “Decoding the Self through Auster’s City of Glass | The New York Trilogy Analysis.” Medium, 8 Sept. 2023, https://adyleho.medium.com/decoding -the-self-through-austers-city-of-glass-the-new-york-trilogy-f09c5c49ed64.

MacIntyre, Alasdair C. *After Virtue: A Study in Moral Theory*. 3rd ed, Bloomsbury, 2011.

Mill, John Stuart, et al. *On Liberty, Utilitarianism, and Other Essays*. New edition, Oxford University Press, 2015. Chapter 5.

Nozick, Robert. *Anarchy, State, and Utopia*. Nachdr., Blackwell, 2012.

Nozick, Robert, *On the Randian Argument*. Wiley, 1971.

Parmenides, Fragments. Est 498 B.C.

"Prime Minister to Create 'Smokefree Generation' by Ending Cigarette Sales to Those Born on or after 1 January 2009." GOV.UK, 4 Oct. 2023,https://www.gov.uk/government/news /prime-minister-to-create-smokefree-generation-by-ending-cigarette-sales-to-those-born-on-or-after-1-january-2009.

Rand, Ayn. "Man's Rights." https://courses.aynrand.org/works/mans-rights/.

Schmidt, Andreas T. "Is There a Human Right to Tobacco Control?" Human Rights and Tobacco Control, edited by Marie Gispen, 2020.

Schopenhauer, Arthur. *The World as Will and Representation*. Dover Pub, 1969.

"The Democratic Party Policies on Substance Abuse." Drug Rehab, https://www.drugrehab.com/ featured/democrats-substance-abuse- policies/.

The Medium of Film: Uncanniness and Narrative Hyper-Realism

Sabina Garcia Ortega

This essay explores the inherent uncanniness of live-action films by analyzing their interplay between concealment and revelation. By utilizing Masahiro Mori's uncanny valley, I argue that certain films can achieve what I label as narrative hyper-realism: the concealment of their contrived nature, embodying human likeness that produces a heightened sense of affinity. I draw on Stanley Cavell's insights into film's foundation and detachment, and Slavoj Žižek's "objet petit a" to understand how film navigates between reality and fantasy. Ultimately, this essay proposes that the medium negotiates between revealing and concealing its uncanniness and that when it successfully conceals it, it achieves narrative hyper-realism. This examination provides a nuanced understanding of the complex relationship between film and its inherent ability to mirror a human perception of reality.

In this essay, I will explore whether the medium of film is inherently uncanny. For this discussion, I will focus solely on live-action films, excluding any form of animation.[146] I will begin by giving an overview of Sigmund Freud's meaning of the uncanny and analyze how it pertains to film, particularly focusing on the interplay between concealment and revelation, where I will suggest that films work to conceal their contrived nature. Drawing on Masahiro Mori's uncanny valley, I will argue that films that succeed in this concealment produce a high sense of human likeness and affinity, occupying the second peak of the graph, while films that

[146] I consider the uncanniness of animation to be an essay in itself: animation's proximity to and imitation of human likeness varies substantially from that of live-action. Animation would most likely fall somewhere between the *first* peak and the uncanny valley of Mori's graph, tracing a different area and movement than live-action.

reveal their contrived nature inevitably fall into the uncanny valley. I will justify the high human likeness and affinity produced by films that conceal their artificiality by drawing on the perspectives of Stanley Cavell and Slavoj Žižek. I will suggest that in such films, the constructed reality seamlessly blends and even surpasses human likeness by presenting a sort of narrative hyper-realism, thus justifying their position in the second peak. The ability of films to either fall into the uncanny valley by revealing their constructed nature, or stand in the second peak by achieving narrative hyper-realism — successfully fulfilling the human desire for a comprehensible reality — reflects both the inherent, but also surpassable, uncanniness of the medium.

In *The Uncanny*, Freud seeks to define what exactly is meant by "uncanny" and identify how the feeling arises. Specifically, he draws on the intricate interplay between concealment and revelation. In Freudian terms, the uncanny is that which was meant to remain concealed but becomes unveiled.[147] I consider this dynamic to be central to the medium of film. As a medium, film partakes in various forms of concealment. To begin, films are composed of sequences of images that quickly change from one to the other, creating the illusion of motion. Images present lifelike objects through their ability to capture 3D elements such as shapes, surfaces, textures, and depths extremely akin to human visual perception. Additionally, in its essence, films are also narrative—they present a story. The combination of these two aspects of the medium then results in realistic objects encapsulated in an artificially constructed manner, in a narrative. However,

[147] Sigmund Freud, *The Uncanny,* 132–3.

and fundamentally so, this constructed nature of film is made to pass unperceived, to depict the narrative as *naturally* flowing. Otherwise, spectators are taken out of the narrative by having the film's momentary resemblance to reality broken. This break of the illusion brings forward the secret the medium of film attempts to conceal, producing an uncanny effect.

Freud's account adds that the uncanny arises when the boundary between fantasy and reality is blurred.[148] I consider this observation to find resonance in the illusion woven by film. The medium of film exists in this liminal space between fantasy and reality as it uses elements from physical reality to create an illusory narrative. Its position in this liminality grants it the potential to become deeply uncanny. When a movie successfully conceals its artificiality, spectators are drawn into the narrative reality, momentarily accepting the constructed world and its logic. However, because the medium of film exists on this border, the uncanny aspect of the medium becomes evident when a film fails to maintain the assumed reality of its invention — film's illusion. This is what tends to happen in what are mostly considered "bad" movies — in these, the medium of film becomes evident. Movies filled with bad acting, an awkward script, clumsy cinematography, and inconsistent storytelling lay bare their artificiality, and the uncanny elements of the medium cease to be concealed. The discomfort produced stems not only from the revelation of artificiality but also from the reminder that what is being witnessed is a carefully crafted attempt at representing reality. In contrast, films that conceal their constructed nature are commonly considered "good" films. Depending on the mastery of the filmmaker, these films

[148] Freud, *The Uncanny,* 150.

produce a reflection that is a "suitably spacious, yet contained, and visually resonant metaphor for the moving images and affective sounds" on the screen.[149] What I will be referring to as "good" films are those that successfully conceal the human hand that carefully produced each second of them, and what I will be referring to as "bad" films are those that (accidentally or purposefully) reveal their assembled nature.[150]

To illustrate the point that for a film to escape the medium's uncanniness it must successfully conceal its constructed narrative reality I would like to point to two specific scenes in David Lynch's *Mulholland Drive*. In the film, there is a scene that occurs twice: Betty's audition. The first time, Betty acts out the script alongside Rita in their house.[151] However, what is particularly interesting about this moment in the film is that it does not work. Spectators can recognize that it is not a real scene within the film — it is an artificial one — it is not part of the reality the movie wishes to create. The scene thus comes across as forged and cheesy. The stilted dialogue and unconvincing delivery exhibit the secret the scene wishes to conceal, momentarily lifting the veil on the fundamentally fabricated nature of the medium of film.

Moments later, when Betty undergoes the real audition, the scene completely shifts — it works.[152] This scene not only completely subverts the audience's previous expectations of how

[149] Bolton, *Contemporary Cinema and the Philosophy of Iris Murdoch*, 27.
[150] To clarify, this is not a critical claim of what makes a movie good or bad. I do not wish to equate "goodness" with concealment—many films that would be considered good reveal their artificiality. Neither do I wish to equate "badness" with an unsuccessful attempt to conceal. Although there is a general pattern that what are considered good movies do not reveal their contrived nature and what are considered bad movies do. I will only use these labels in the broad sense I have outlined for clarity and conciseness.
[151] Lynch, *Mulholland Drive*, 01:10:24–01:11:34.
[152] Ibid*,* 01:17:33–01:20:37.

the audition will go, but also brings about a deeply uncanny feeling, as this second reiteration is inevitably compared with the first. It is made obvious how the dynamic between the actors, the camera work, and the delivery of the lines, change in *exactly* the necessary way so that everything is positioned to imitate the human perception of reality convincingly. Through these changes, audiences become momentarily convinced and absorbed into the events taking place in this instance.

The first iteration of the audition scene (before being explicitly revealed as Betty and Rita rehearsing a script) is itself an uncanny moment — it presents a "bad" scene in the movie where audiences are reminded they are watching a movie. Something that was premised as being real within the movie is revealed to be orchestrated. However, the second iteration further underscores the uncanniness of the medium, as the dialogue that had already been established as constructed is made to momentarily feel real precisely because it once again hides its contrived nature through the cinematic technique — it is made "good." This second execution becomes uncanny because it makes obvious the illusion of the medium of film by emphasizing what was not well-executed before. This careful interplay makes obvious the constructed narrative artificiality of the medium of film. It shows how, under correct execution, film presents narratives in a way that seems real, so that we momentarily forget that they are narratives. Thus, the conjunction of these two extremely similar but enormously different scenes reveals how the medium of film possesses a great uncanny potential.

I believe that film's existence in this peculiar position as a medium that can both reveal and conceal its uncanniness can be explained through the roboticist Masahiro Mori's "uncanny valley," which seeks to graph a particular realm of human perception and affinity. As seen in Fig. 1., Mori graphs the level of affinity felt for an entity against its level of human likeness. The line delineates Mori's proposed trajectory. He suggests that, as non-human entities approach human likeness, the affinity increases, until it reaches a critical point. At this point, the sense of affinity rapidly begins decreasing, until it plunges into negative affinity.[153] This is what Mori labels the uncanny valley. The uncanny valley traces this space characterized by a sudden negative affinity, invoking an eerie sense of strangeness and aversion.[154] Yet, Mori proposes that, as the entities continue to progress in human likeness, the affinity ascends once again, resulting in an even higher peak than before.[155]

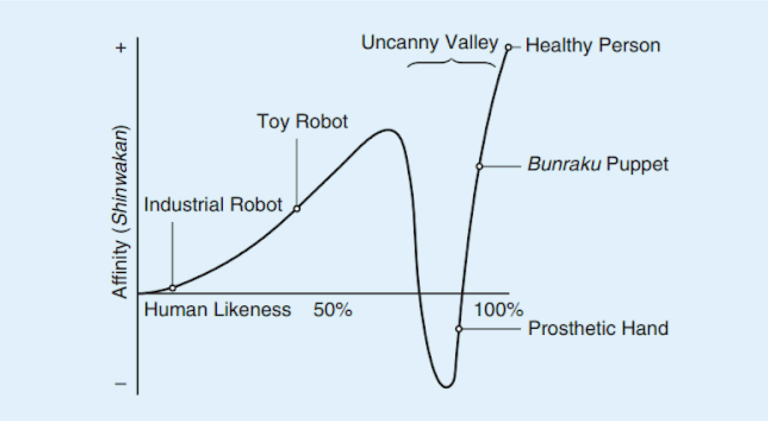

Fig.1. Mori, "The Uncanny Valley," 99.

[153] Mori gives the example of not realizing a person's limb is prosthetic until one touches it and senses it to be cold.
[154] Mori, "The Uncanny Valley," 99.
[155] Ibid. As an additional note, Mori also adds that movement intensifies both peaks and the valley (99).

I propose employing this nuanced movement between uncanniness and affinity as a theoretical framework for exploring the uncanny aspects inherent to the medium of film. Mori's theory provides insight into how the medium of film can navigate the spectrum between uncanniness and affinity either by revealing its contrived nature or effectively concealing it. In essence, "good" films (those that conceal their artificiality) occupy the second peak (the highest sense of human likeness and affinity), while "bad" movies (those that reveal their constructed nature) fall into the uncanny valley (a strong sense of human likeness but a negative affinity). This showcases how the medium of film lends itself to revealing its uncanny disconnect from reality. "Bad" films just tend to reveal both their attempt at imminent reality and their complete disconnect from it, more often unintentionally. The medium of film then can mediate between the two spaces in Mori's graph. It can either descend into the uncanny valley, as is the case with "bad" movies when revealing their artificiality, or stand on the second peak, offering a sense of indistinguishable human likeness by creating a narrative representation of reality, a concept I term as narrative hyper-realism.

The juxtaposition between the acting scenes in *Mulholland Drive* encapsulates the dynamic interplay between the uncanny valley and the second peak of the medium of film. I contend that the initial revelation of the constructed nature induces a temporary sense of strangeness, having the scene fall into the uncanny valley. However, during the actual audition, the scene ascends to a heightened state of narrative hyper-realism through its convincing delivery that conceals the artificiality of the medium.

In *The World Viewed*, Stanley Cavell delves into key elements that shed light on how the medium of film generates its narrative hyper-realism. According to Cavell, the foundation of the medium of film lies in the succession of photographs whose placement captures an automatic projection of the world.[156] These sequential images maintain a sense of presentness within the depicted reality, giving the impression that events are unfolding. Simultaneously, the audience acknowledges their physical absence from these events.[157]

Further, we should note that the medium of film not only distances the spectators from events, as Cavell observes, and conceals the human creator, but the medium *hyperbolizes* the absence of the human creator. "Good" movies convince us that there is no human creator orchestrating the events taking place, and the narrative hyper-realism passes unperceived, occupying the second peak. The medium's automatic hiding, cutting, and framing presents reality in a digestible way, reinforcing the narrative hyper-realism. Conversely, "bad" films precisely remind audiences of this, revealing the secret that should have been kept hidden — that all of what is being depicted is false — thus, plunging into the uncanny valley.

I consider that the combination of these two elements contributes significantly to creating the narrative hyper-realism of the medium of film. The sense of presentness creates just a sufficiently absorbing experience while the spectator's inevitable detachment — absence from the events taking place — allows for the feeling that the reality presented — although a carefully

[156] Cavell, *The World Viewed: Reflections on the Ontology of Film*, 16, 72–3.

[157] Ibid, 22–3, 25.

constructed and orchestrated one — is more tangible as it is more comprehensible than the often perplexing nature of actual reality. Film's medium precisely allows the spectator to uncover its reality, a reality they will never form part of, but one that is central for it to be perceived and created.

To exemplify what I mean by narrative hyper-realism I would like to allude to Christopher Nolan's film *Memento*. The film intricately manipulates time and memory, unfolding its narrative in a non-linear fashion. This nonlinearity immerses spectators in a mental state that mirrors that of the protagonist, Leonard. However, and very crucially so, the revelation of the narrative's key element does not take place until the end of the movie.[158] *Memento*'s ending renders its narrative comprehensible — it allows the movie to make sense. However, it becomes comprehensible only to audiences (it's a matter of seconds before Leonard inevitably forgets once again). In the film's ending, the audience and possibly Leonard (momentarily) escape this confusion.

I consider this delayed revelation to be more than just a plot twist; it encapsulates the essence of the narrative hyper-realism the medium of film can achieve; the accurate sequence of Leonard's story is finally *made* comprehensible exclusively for the audience — each second in the film is placed into a comprehensible order. Through this narrative hyper-realism, *Memento* imbues coherence into its preceding complexity. Prior to this revelation, audiences occupied a similar position to Leonard, navigating a reality that was simultaneously familiar and perplexing.

[158] Nolan, *Memento*, 01:43:26–01:48:35.

By meticulously portraying a humanly constructed reality, *Memento* allows its spectator to become the sole understander of the film's reality. This comprehension arises precisely because the narrative, as a story, is fundamentally graspable and framed. This is what I refer to as narrative hyper-realism.

The comprehensible narrative presented by film represents a reality that can be gripped. In everyday life, we tend to be Leonard, struggling to sustain any deep understanding of our reality — whether due to reluctance or to the inherent limitations of our existence (in the case of Leonard, this is illustrated by his inability to form new memories). The narrative, therefore, becomes hyper-realistic by providing the comprehensible reality we yearn for, or constantly wish to deceive ourselves that we obtain, establishing a profound human psychological likeness. This is what allows films that surpass the inherent uncanniness of the medium to occupy the second peak and result in such a great sense of affinity.

An immediate counterargument to my claim arises with surrealist movies. These movies, characterized by their non-linear narratives and dreamlike sequences, challenge the conventional understanding of films. Surrealist cinema, by its very nature, disrupts the natural world projections associated with narrative hyper-realism. Instead of offering a comprehensible narrative, these films plunge viewers into a realm where logic and continuity are abandoned. To this, I reply that the incomprehensibility of surrealist films is a statement in itself. The comprehensible message of reality that surrealist films seek to present is that we cannot make sense of reality. Thus, the medium of film, as a carefully constructed projection of a fragment of reality, inevitably delivers a narrative (even if this narrative is that there is none).

Finally, I would like to explore Žižek's concept of "objet petit a." For Žižek, the objet petit a is the trace of the real, perpetually perceived in a distorted way. It embodies the surplus of confusion and disturbance arising from the pursuit of an objective reality. In the context of film, this distortion becomes a fundamental element, as it plays a crucial role in the medium's negotiation between reality and fantasy. Žižek contends that the objet petit a is always perceived in a distorted manner because, at its core, it does not exist outside of this distortion, outside of our own, inevitably flawed perception of the real.[159] The medium of film, acting as a conveyor of a fantasy deeply entwined with reality, inherently distorts reality. This distortion takes on an extremely familiar form — a narrative one. However, the extent to which this distortion passes unperceived determines whether a film occupies the peak of the graph or descends into the uncanny valley.

The objet petit a applied to film encapsulates the essence of narrative hyper-realism seen in "good" movies. The distortion introduced by the desire for a digestible reality, manifested through the deliberate construction of narrative and visual constructions, contributes to the immersive quality of cinema. What are generally considered "good" movies are thus those that excel in hiding that they are presenting a distortion and in creating a reality that resonates with viewers' desires and expectations. Conversely, "bad" movies are therefore those that remind

[159] Žižek, *Looking Awry: An Introduction to Jacques Lacan through Popular Culture*, 10, 49.

spectators of the artificiality inherent in the medium of film. Žižek draws on Lacan's point de capiton, the point where a situation perceived as natural or familiar becomes denatured, uncanny, when a detail that does not belong, that seems odd, is revealed.[160] I consider this to be precisely what happens in "bad" movies — those that remind spectators of the artificiality inherent in the medium of film. They pull spectators out of this constructed reality by reminding them that they are watching a human representation of a conceived reality.

Žižek's argument extends to the illusion of narrative flow. Narratives, despite their apparent coherence, conceal the retroactive nature of their consistency. The ending, retroactively assigning meaning to preceding events (as can be seen in the case of *Memento*), exemplifies the manipulation of desire and distortion. It conceals the fact that at every point, things could have gone in a different direction. The concealment of its artificial construction precisely allows for the narrative depicted to be taken as natural and originally flowing, without any type of external intervention.[161] Like the objet petit a, the film is perceived in a distorted way, maintaining an illusion of narrative flow while concealing the external interventions that shape its coherence. This process often goes undetected, embodying a great sense of human likeness as it mirrors a reality through human perception that captures just enough of actual reality to feel genuine. The narrative coherence, retroactively imposed, satisfies the viewer's desire for a comprehensible experience. This is precisely the conceit of the medium of film, the creation of a reality that

[160] Ibid, 55.
[161] Ibid, 40.

appears to spectators as if it emerged organically, mirroring the human perception of our reality. It is a reality that appears so close to the actual reality that we are willing to accept it — even more than actual reality because it is more comprehensible.

In conclusion, the medium of film's uncanniness is not solely rooted in its implication of physical connection with objects but is equally embedded in its meticulously crafted, yet often concealed, nature. Positioned on the boundary between reality and illusion, film operates in a unique space that can either expose or effectively hide its inherent uncanniness. Mori’s graph, illustrating the relationship between human likeness and affinity, provides a valuable framework for understanding how film oscillates between revealing and concealing its uncanny aspects. “Good” movies, those that do not reveal that they are a carefully meditated and created narrative, occupy the highest sense of human likeness and affinity. These films achieve narrative hyper-realism by presenting a carefully constructed representation of reality that resonates with the spectator’s desires for a comprehensible reality, offering the illusion of gripping the trace of the real. On the other hand, “bad” movies, by exposing their contrived nature, diminish their human resemblance. The acknowledgment of their contrived narrative breaks the illusion and, akin to Lacan’s point de capiton, brings forward the oddity (the artificiality), propelling these films into the uncanny valley.

Works Cited

Bolton, Lucy. *Contemporary Cinema and the Philosophy of Iris Murdoch*. Edinburgh: Edinburgh University Press, 2019.

Cavell, Stanley. *The World Viewed: Reflections on the Ontology of Film*. Cambridge, MA: Harvard University Press, 1979.

Freud, Sigmund. *The Uncanny*. Translated by David McLintock. London: Penguin, 2003.

Nolan, Christopher, dir. *Memento*. Santa Monica, CA: Summit Entertainment, 2000.

Mori, Masahiro. "The Uncanny Valley." Translated by Karl F. MacDorman and Norri Kageki. *IEEE Robotics & Automation Magazine* 7, no. 4 (1970): 98–100.

Lynch, David, dir. *Mulholland Drive*. France: Les Films Alain Sarde, 2001.

Žižek, Slavoj. *Looking Awry: An Introduction to Jacques Lacan through Popular Culture*. Cambridge, MA: The MIT Press, 1992.

The Multiplicity of Oppression: Young's Five Faces Explored Through Luke's Dimensions of Power

Nika Evenson

This paper critically analyzes Iris Young's evaluative framework of oppression in conversation with Steven Lukes' three-dimensional power philosophy. Young's approach, centered around the Five Faces of Oppression and the recognition of systemic constraints, represents a departure from traditional notions of overt tyranny and domination. By emphasizing structural phenomena, she brings attention to the hidden and insidious aspects of oppression often overlooked in our awareness.

However, this paper argues that Young's framework, while valuable, has limitations in its rigid categorization, which appears through its use of structural phenomena. The introduction of Steven Lukes's Dimensional Powers offers an alternative perspective that accommodates the fluid and dynamic nature of oppression. Lukes's three dimensions — overt power, shaping political discourse, and subtle influence — provide a nuanced understanding of varying levels of oppression and account for individual experience.

The analysis suggests that Lukes's dimensional power approach may offer a more comprehensive and adaptable framework for understanding oppression. It allows individuals to pinpoint where and how they experience oppression and recognizes the importance of addressing covert forms that influence beliefs. While Young's framework is accessible, Lukes's perspective provides a greater exploration of oppression's complexity, encouraging a more just and equitable society by addressing diverse experiences of oppression. In conclusion, both contribute valuable insights, but Luke's dimensional power approach appears more comprehensive for understanding and addressing oppression in society.

1. Introduction

Iris Young's Five Faces of Oppression provides a contemporary understanding of oppression that transcends previously conceived notions of its kind within political philosophy. Following a demonstration of the discrepancies within methodological individualism[162] and structural phenomena[163], she provides several descriptive forms that a collective might experience when oppressed.[164] The five forms — exploitation, marginalization, powerlessness, cultural imperialism, and violence — work as a foundation for the classification of oppression under her framework. Through an evaluation of her reasoning, I argue that in analyzing oppression using structural phenomena, as Young does, we risk fostering a framework that does not fully reflect the varying levels of oppression within society. This perspective is substantiated through an exploration of Stephen Lukes's Dimensional Powers, which I believe provides a more comprehensive understanding of oppression without attempting to fit a multiplicity of experiences into a set of descriptive forms. I claim that Luke supplies a framework where oppression can fluctuate in severity according to individual or collective experience through the employment of systematic levels. Furthermore, Lukes observes oppression through a lens that does not focus solely on structural phenomena, opening itself to a larger variety of events. When put into conversation with Young's faces, Young's framework starts to function instead as more

[162] Methodological individualism provides a framework for understanding social phenomena that occurs through an exploration of individuals that incites society as the outcome of their actions and intentions.
[163] Structural phenomenon, according to Young, refers to institutional rules or regulations that immobilize or diminish a social group.
[164] Young, "Five Faces of Oppression," 40.

accessible examples of Lukes's dimensions, aiding individuals to better understand the circumstances and effects of their oppression and not as a framework in and of itself.

Young begins her paper with an explanation regarding the definition of oppression, discussing how the term "oppression" has been reinterpreted to encompass more phenomena within the last century. Originally, the term was used to describe events such as apartheid where tyranny and domination were blatantly visible, leading many individuals to believe that oppression is no longer relevant in our society.[165] Young discusses how these individuals view oppression as something that would be inflicted upon them by an outside entity, such as a foreign power and not by their community or government. Within much of western society there are no outside powers that dominate its citizens or their rights and freedoms; therefore, they are theoretically free of oppression. However, many are forced to acknowledge the active nature of oppression within their society through personal experiences or second-hand accounts. Whether due to religion, ethnicity, or gender, Young suggests that oppression should be understood as systemic constraints on various groups.[166] This leads oppression to be entrenched in the structural foundations of many institutional rules, and as a result, the norms, beliefs, and values of those that follow them.[167] Individuals who perpetrate this type of structurally embedded oppression often do not see themselves as agents of oppression and are instead unaware of the harm they may inflict upon other groups.

[165] Ibid, 40.
[166] Ibid, 41.
[167] Ibid.

Before defining her five categories of oppression, Young first describes what makes an individual inherently part of a social group (thus applicable for evaluation), a term which she differentiates from an aggregate or association. An aggregate is a simple classification that relates to a visible attribute such as eye color, but also gender, skin color, and age. The phenomenon that separates an aggregate from a social group is that in an aggregate, the individual exists prior to the collective. In other words, the classifications are not a necessary part of their identity. For example, the classification of an individual through an external or accidental attribute like eye color, would not reflect their internal disposition, personality, or outlook on society.[168] On the other hand, an association is understood as a "…formally organized institution…"[169] which would entail voluntary participation in entities such as clubs, corporations, political parties, or churches. Through the lens of these two terms, Young defines social groups by their direct connection to the identity of the individual, which distinguishes them from other collectives due to culture, religion, or way of life. Furthermore, social groups and their identities exist "…in the encounter and interaction between social collectives,"[170] more specifically, they exist due to the differences between individuals who consider themselves a part of the same society. These definitions are necessary because they outline the reasons why multiple groups can be evaluated as oppressed while existing within the same public sphere. In

[168] Ibid, 44.
[169] Ibid.
[170] Ibid, 43.

looking at other ways of grouping individuals, there is a noticeable variation in the effect of separation due to identity and its importance to who a person is and not simply how they appear.

The forms of oppression that Young has defined within her Five Faces of Oppression are exploitation, marginalization, powerlessness, cultural imperialism, and violence. The foundation of exploitation is that oppression occurs when one social group benefits from the labor of another social group, which has been steadily transferred to them over a period of time.[171] This is employed structurally through a systematic process that is consistently maintained in order to ensure the power, status and wealth of the benefitting social group.[172] The next form, marginalization, is understood as a deprivation of material items through distribution injustice while also implying a "deprivation of cultural, practical, and institutionalized conditions"[173] that do not allow the marginal to utilize their capacities to achieve recognition and interaction. Powerlessness defines those who lack authority, particularly in the division of labor, which results in them having to take orders without any creative or meditative autonomy.[174] This form of oppression is easily visible in capitalist countries, such as the United States, where workplaces function under hierarchical systems that do not allow most individuals to contribute in decision making.[175] The fourth form of oppression that Young describes is cultural imperialism, which is defined as a universalization of a dominant group's culture and its establishment as the norm for

[171] Ibid, 49.
[172] Ibid, 50.
[173] Ibid, 55.
[174] Ibid, 56.
[175] Ibid.

all groups within that society.[176] The final form of oppression is violence, which includes "harassment, intimidation or ridicule simply for the purpose of degrading, humiliating or stigmatizing group members."[177] The reason that Young provides this as a notion of oppression is due to the structural and social setting that has allowed violence to go unchecked and in some cases be found acceptable in society.

Lukes' framework highlights levels of power that do not rely on the various descriptions of oppression Young defines within her forms. In this regard, Young's faces of oppression allow Luke's dimensions the possibility to be explored through accessible definitions, such as cultural imperialism, which provide groups and individuals a starting point to explore their oppression. Beginning with one-dimensional power, defined as "overt power"[178] meaning that this power is observable and mainly related to active political agents and organizations, which may refer to 'violence'. Two-dimensional power is then understood as power that shapes the political sphere by deciding what can and cannot be discussed by indirectly influencing what options may even be considered structurally, potentially encompassing 'powerlessness'. Finally, three-dimensional power is noticeably more latent, as its power operates by defining people's interests primarily by subtly manipulating an individual's value system and beliefs, as illustrated by Young's concept of 'cultural imperialism.' Furthermore, due to its nature as a covert conflict, many individuals who have succumbed to its influence are unaware that their preferences have been shifted.

[176] Ibid, 59.
[177] Ibid, 61.
[178] Lukes, *Power: A Radical View*, 4.

Through an observation of Lukes's framework, we can see there are areas of overlap within Young's classes, particularly in how each criterion can be encapsulated within one of Lukes's dimensions. For example, exploitation, which focuses on the unlawful utilization of another's physical or metaphysical labor, can be considered one-dimensional due to the direct presence of power over a group, two-dimensional if it pertains to exploitative rules and regulations, or three-dimensional if the subjugated individual is unaware of their exploitation. Moreover, a category such as powerlessness could be considered primarily two-dimensional as it is typically systematic, as found within regulations and guidelines that aim to restrict the power of another group. However, powerlessness can pertain to the first and third levels as well, should an individual be defenseless against another's actions or theoretically to the point in which they are conditioned to accept their position. This is different from marginalization, as this form of powerlessness is likely to occur in a smaller setting, potentially a romantic relationship where an individual is unable to change their circumstances. This individual is still oppressed as any social group may be, but its classification under Young would be overlooked due to its singular nature regardless of how many individuals suffer from the same situation. Violence is another example, as under Young's description, it fits largely into one-dimensional power. Nevertheless, a closer examination can isolate violence into at least three separate areas that highlight its physical, manipulative, or psychological nature. Physical violence would still likely fall under the first dimension, as it is visible power exerted over another individual, though if it was allowed due to manipulative tactics such as propaganda or gaslighting, it could be considered within the second dimension. Psychological violence would likely be entrenched in a value or belief system, such as religion where their belief in their god is tied to their acceptance of violence against them.

Each facet of oppression that Young presents can fit into Luke's already diverse system of dimensional power that does not necessitate categorization. This is one reason it may provide a more illuminating way to analyze oppression. For instance, Young claims that "applying these five criteria to the situation of groups makes it possible to compare oppressions without reducing them to a common essence or claiming that one is more fundamental than another."[179] However, as I discussed previously, each criteria has an inclination for a certain dimension of power. Young's notion evidently does not imply fundamentality or commonality, but nonetheless, dimensional powers are able to provide a more in-depth and cohesive understanding of oppression, where further interpretation can occur. They are not simply providing an area for the evaluation of oppression under descriptive terms in a large-scale society as Young does, but they truly create the evaluation and exploration of oppression in a fluid manner. Lukes establishes a notion of varying levels of oppression that coincide with an individual's overall freedom in its most specific forms. In other words, Lukes puts into words how even isolated cases of oppression are brought about without relying on specific descriptive criteria. Moreso, Young utilizes descriptors such as violence and exploitation, and requires that the social groups fit into one of forms to be evaluated as oppressed. On the other hand, Lukes employs open and

[179] Iris Marion Young, "Five Faces of Oppression," 64.

non-specific criteria in his framework as he is discussing forms of power, which only implies that power of some type must be exerted.

In creating these categories, Young creates boundaries between the various types of oppression experienced by collectives, while leaving isolated cases overlooked. She discusses the categories as if they are “multiple, cross-cutting, fluid and shifting”[180] but discusses the requirement of one of her Five Faces of Oppression for a social group to be evaluated as oppressed. She explains how group differentiation is not necessarily oppressive[181] but does so by providing the decline of parochial attachments[182] as a reason for the position, which I believe to be particularly outdated. In this context, I believe Young is implying that group differences are not as inherently oppressive in present-day society due to globalization. As individuals are less likely to be confined to small communities where being perceived as ‘different’ may have resulted in them being oppressed. Furthermore, Young considers how markets and social administration have caused an increased global social interdependency[183], whereas I believe she does not consider the popularity of social media, which has transcended traditionally aggregated social groups and boundaries. These social groups formed with the help of social media can no longer be defined by external attributes or location because they cannot be found in one specific society. Social media has led to an interconnectedness where a group such as this does not

[180] Ibid, 48.
[181] Ibid, 47.
[182] Ibid.
[183] Ibid.

interact within any systemic space that would lead to them being oppressed under Young's forms. I make this argument because if a group is marginalized on a social media platform due to their values or culture, they are likely experiencing real-world marginalization as well. Nonetheless, the systemic results of oppression vary considering the location of a group due to the institutions, laws, and regulations that are applicable. Therefore, if social media is used to instigate one of Young's forms of oppression in a circumstance that does not already pertain to any real-world instances of oppression for that individual or group, then it cannot be evaluated through the framework. Consequently, even if social media falls under the descriptive category of oppression presented by Young, it cannot be contained within structural phenomena and systematic setting that Young has set as her foundation. I do, however, acknowledge that Young formed this notion before the conception of modern social media, making it significantly more applicable in the past. Nevertheless, I provide this as an example of how a Lukesian three-dimensional approach to oppression provides a more universally relevant form of evaluation.

Lukes's approach is able to observe occurrences, such as social media, on a variety of nuanced levels, as the primary focus of his work is power. Power has yet to be defined under a single description, as Lukes suggests power can be polysemic, meaning that its definition shifts according to what is most appropriate in that context.[184] Power could fall under what Wittgenstein refers to as a 'family resemblance,' implying that it has no common substance, or

[184] Lukes, *Power: A Radical View*, 61.

potentially power is dependent on local 'language games.'[185] In any debate, argument or dialogue, power could be employed and understood in a multitude of ways according to the desires of the individual, the location, political beliefs, age, gender, what have you. Therefore, Lukes explores power as a 'dispositional concept', which entails a "conjunction of conditional or hypothetical statement[s]" that identifies the possible situations in which power is or could be employed.[186] In this regard, when Young utilizes structural phenomena as a foundation for her five faces, she limits her forms of oppression to a structurally physical environment where institutions and social groups interact under the same regulations. Conversely, Lukes's notion of power operates using the "abilit[ies] or capacit[ies] of an agent or agents," regardless of whether they actively use these capacities.[187] When considering social media's immaterial nature, a systematic approach of oppression cannot be implemented on a global level, at the very least not currently, whereas a dispositional power approach is able to account for the various differences across societies and continents.

Lukes's dimensions of power can provide an individual with the understanding of how and where within a system they are being oppressed, while also providing insight into what properties are being affected. For example, by employing powerlessness on a religious minority, they would theoretically be excluded from decision-making. Whereas, the dominant religious social group would be able to create laws and regulations that could make it easier for them to

[185] Ibid, 61-2.
[186] Ibid, 63.
[187] Ibid.

practice, wear their religious symbols openly, express their views while making it more difficult for the religious minority to do so. Eventually, the religious minority becomes aware of their powerlessness; however, they have been effectively marginalized and are now dependent on the state. The religious minority is able to evaluate their situation by exploring oppression through powerlessness and marginalization within Young's framework. However, it is increasingly unlikely that any minority is simply oppressed through one or two systemic factors as the religious minority had been. Oppression does not only exist in structural phenomena, as there are individual, ideological, and social factors that are necessary to consider. However, they are not encapsulated within the systematic basis of Young's Faces of Oppression. Young's goal was to evaluate oppression on an institutional level which is meant to be accessible to social groups while not providing a framework that ranks the forms of oppression experienced by these groups. This reality does not lessen the usefulness of Young's forms; it simply highlights areas in which improvement and innovation is necessary. By combining Lukes's and Young's approaches to power and oppression, there is an opportunity to implicate structural oppression and dimensional powers within the same framework. This approach would be able to encompass the common forms of oppression that Young presents, while being able to observe and evaluate oppression on a smaller scale than social groups and in a non-structural setting. Lukes's dimensional powers were shown to be able to categorize oppression that occurs on social media, while Young's failed to do so. However, Lukesian methodology is difficult to understand without explicit knowledge of one's situation. Therefore, through an amalgamation of both works in which Lukes's dimensional powers provide the foundation of the framework and Young's faces of oppression provide accessible explanations and descriptive criteria, it is likely that a more globally applicable structure for evaluating and explaining oppression may be found.

Iris Young's evaluative framework of oppression in dialogue with Steven Lukes' three-dimensional power philosophy provides further insight into the complex nature of negative societal constraints. Young's notion of oppression is understood primarily as structural constraints through her Five Faces of Oppression, which marks a significant departure from traditional notions of tyranny and domination. Her focus on structural phenomena brings awareness to the hidden and insidious characteristics of oppression that often elude our awareness. Lukes's three dimensions of power offers an alternative approach that more easily accommodates the fluid and dynamic nature of oppression through the introduction of a more systematic approach. The notion of dimensional power establishes the opportunity for a more nuanced and flexible understanding of oppression, allowing individuals to pinpoint where and how they are oppressed and what aspects of their freedom are directly affected outside of social groups. It underscores the importance of recognizing and addressing covert forms of oppression, which often render large collectives' agents unaware of their own subjugation. However, it is likely that Young can provide a more easily accessible approach to oppression, as Luke's framework may be difficult to understand if an individual is not actively aware of their society and their relation to different social groups. In this case, Young's Five Faces of Oppression may be used to explore oppression and where it may fall within Lukes's framework.

In essence, while both Young's and Lukes's perspectives offer valuable insights into oppression, the discussion here suggests that the dimensional power approach might provide a more comprehensive and adaptable framework for understanding and addressing the diverse experiences of oppression in a society. It encourages us to not only be aware of large-scale oppression and the overt constraints of a system but also individual experience and the subtler influences that shape our preferences and beliefs, ultimately moving us closer to a more just and equitable society.

Works Cited

Lukes, Stephen. *Power: A Radical View*, 2nd Ed. (Macmillan 2005), 14-107.

Young, Iris Marion. “Five Faces of Oppression,” The Philosophical Forum 19,4 (1988), 39-65.