APXH

ARCHÉ

JOURNAL OF PHILOSOPHY

Arché

SPONSORS

PLATINUM

Highland House Boutique Hotel

Glenn and Debbie Street

GOLD

Gillian Howcroft

Geri Jones

Ted Stinson

Luanne Thorndyke

ACKNOWLEDGMENTS

CONTENTS

ABOUT THE CONTRIBUTORS

REBECCA AMEN *recently graduated from Middlebury College with a B.A. in English & Philosophy. Rebecca's interests include moral philosophy, the philosophy of human rights, and the philosophy of language, and she hopes to ultimately bring these interests to the study of law.*

MISHA BELSKY *is a M.Phil. student at Cambridge University, focusing on problems in contemporary metaphysics. In his first term, Misha developed Kit Fine's notion of global indeterminacy to resist the Argument From Vagueness (for mereological universalism). In his second term, Misha presented a novel defense of simplicity — or, less simply, of 'qualitative ontological parsimony' — as an epistemic virtue in first-order metaphysics. For his final thesis, Misha entered the philosophical debate over time's arrow (the B-Theory vs. the C-Theory of time) centered around the Past Hypothesis. Misha critiqued several of Tim Maudlin's defenses of a primitive temporal asymmetry in the world's fundamental structure, and found the C-Theory's (anti-realism about time direction) extant appeal to parsimony -- as the tie-breaking advantage over realism -- to be underdeveloped.*

As a retired veteran, ARNOLD DINH *decided to study overseas in New Zealand to optimize his military benefits. His benefits allowed him to play rugby at the college level and earn his B.A. in Philosophy.*

BRIGHT DUA-ANSAH *is a master's student at the University of Tartu (Estonia), and formerly a Philosophy Teaching Assistant at the University of Ghana where he completed his undergraduate studies. He is keen on student-led clubs on bridging faculty-student gaps in aforementioned universities. His professional background, however, is in qualitative research on public policy and child safety & health. He works in Fintech and is an aviation and boxing enthusiast.*

KYLIE FELICIANO *is graduating with a major in Philosophy & Neuroscience and a minor in Innovation & Entrepreneurship. Her interests are in ethical technology development and investment, and she hopes to pursue a career in law upon graduation. In her free time she enjoys painting and reading dystopian and science fiction novels, her favorite being Mary Shelley's Frankenstein.*

SANTIAGO FLÓREZ SÁNCHEZ *is completing his master's degree in Philosophy at the University of Los Andes (Bogotá, Colombia). His primary research interests include philosophy of mind, psychology, and neuroscience.*

RASHAD REHMAN *is a Ph.D. student in the Department of Philosophy at the University of Toronto, as well as a Collaborative Specialist in Bioethics at the Joint Center for Bioethics.*

EVERET SMITH *is a Ph.D. candidate at Emory University. They received their B.A. in Philosophy at Appalachian State University in 2018, and their M.A. in Ethics and Applied Philosophy at UNC Charlotte in 2021. Their research interests include Frankfurt School critical theory, dialectical and historical materialism, Latin American philosophy, Chicana feminism, psychoanalysis, and phenomenology, among others. Their publications may be found in such places as Aztlán: A Journal for Chicano Studies (August 2023), Ethics and Information Technology (July 2022), and the American Society for Aesthetics Graduate Journal (December 2021).*

LIOR TAPNACK *(B.A.) is a recent Philosophy graduate from Durham University. Despite the metaphysical nature of his paper for Arché, Lior spent most of his undergraduate degree focused on practical ethics. His undergraduate thesis, titled "Expecting Everything and Achieving Nothing — Responding to the Recklessness of Longtermism," is an attempt to disprove the probabilistic consequentialism of 'longtermism,' a controversial moral theory gaining traction amongst the effective altruism community. Beyond Philosophy, Lior is passionate about disruptive technology and is currently a strategy consultant working with growth-stage technology companies headquartered in Israel.*

BY ANY STRETCH OF THE (MORAL) IMAGINATION:
A DEFENSE OF CONTEMPORARY POETRY'S
ANIMATION OF MORAL ACTIVITY

by Rebecca Amen
Middlebury College

ABSTRACT ‖ *In this paper, I first explore literature's connection to moral philosophy in the context of Aristotelian philosophy and Martha Nussbaum's writings on the matter. I pay specific attention to Nussbaum's claim that the highly particular, morally-attuned novel best expands the "moral imagination," thus enhancing everyday moral behaviors. Then, I investigate the features of a novel claimed to contribute to these behaviors, subsequently contrasting these with the features found in many contemporary poems. Finally, I argue that these opposite features found in such poems build the same moral behaviors to which Nussbaum refers in equally valuable ways, suggesting that Nussbaum's paradigm of moral literature can be expanded.*

I. INTRODUCTION

As they confront the unpredictable moral challenges that arise out of our interactions with others, individuals who want to behave with optimized consideration and compassion ought to draw on all of their prior knowledge and experience to do so. They do not want to see the circumstance of the person before them as informing their future moral behavior. Rather, they want a way to broach this present interaction preparedly and sensitively—a way to treat their counterpart's situation as an end in itself. How do we study the maximization of ethical behavior in our interactions with others? Perhaps

only the moral philosophy student does so explicitly. The rest of us may implicitly pick up moral lessons from our own experiences. However, if we aim to avoid reducing our past interactions to moral lessons, we can also look to the templates or suggestions of behaviors that representations of life lay out. Here, we turn to this paper's particular concern: the connection between literature and everyday moral activity.

There are, of course, many antecedent views on this connection, the recognized relationship itself dating back to the paradigmatic "ancient quarrel" between Plato and Aristotle about the moral value of Greek tragedy (Plato 2016). This debate in Western philosophy is ongoing, and it has been renewed recently in the writings of Martha Nussbaum, who conceives of the elaborate, morally-attuned novel as the ideal form of literature for moral education. Indeed, Nussbaum's views draw heavily upon an Aristotelian conception of tragedy as the best genre to provoke moral deliberations. Since the tragedy of classical Greek antiquity and the novel of modernity share much in common, both forms of literature may work to strengthen what Nussbaum refers to as our "moral imaginations." We can think of the moral imagination as what informs our moral behaviors, allowing us to imagine the particular situation of our counterpart and to react with sensitivity. In this way, the moral imagination improves our treatment of the Other. I will demonstrate in what follows that certain genres of literature are useful informants and expanders of the moral imagination, and I will use this idea of employing the moral imagination as a conception of everyday morality. I will, however, disagree with the claims that place complex and narrative forms of literature as *the* ideal for moral education. I will show that certain literature of that ilk with vastly different formal characteristics—namely, contemporary poetry—does something unique and equally valuable for broadening the moral imagination and optimizing its uses. Before demonstrating this opposition of characteristics and illuminating what each trait of contemporary poetry does for the moral imagination, I will spell out Nussbaum's argument for the novel as a paradigmatic genre of moral literature.

II. NUSSBAUM'S IDEAL MORAL LITERATURE

In her book *Love's Knowledge,* Nussbaum draws a connection between literature and moral philosophy by positing that forms of literature expand the moral imagination. To demonstrate this, Nussbaum strives to locate a literary piece that can simultaneously function as a work of art and as moral philosophy to argue how moral education moves toward active and practical improvement of moral behaviors. In what follows, I will first clarify the term "moral imagination"; it will be an ethics of everyday interactions that is different from deontological or consequentialist moral systems . I will then touch on how Nussbaum's idea of this kind of novel as an ideal genre of moral literature is derived from Aristotle's views of tragedy and morality. Subsequently, I will describe the main features that distinguish Nussbaum's paradigmatic moral literature.

1. The moral imagination and Nussbaum's Aristotelian conception of morality

To begin, a clear definition of the term "moral imagination" is in order. A person uses her moral imagination when she projects herself into another's context, envisioning herself in that other person's shoes. Here, "context" should be understood in the broadest sense, encompassing all of an individual's experiences. We may closely identify with this context, using our moral imaginations to approximate the finer-grained subjectivities or to locate the similarities in another's experience with our own, but we may find the context foreign. Unfamiliar terrain most strenuously exercises the moral imagination. The moral imagination, as such, is a resourceful thing; it can conceivably make use of all experience, both firsthand and "vicarious" (as through literature or other forms of representation), and of all knowledge. Nussbaum did not originate the concept of the moral imagination; in fact, Edmund Burke is thought to have created the concept (Burke 2009). However, in this paper we will focus on the ethics that Nussbaum builds around the moral imagination.

Nussbaum emphasizes the improvisational quality of actions arising from the use of moral imagination, juxtaposing this kind of alternative ethics to the systematized conceptions taught in moral philosophy classes (Nussbaum 1990). The type of ethics resulting from the usage of the moral imagination centers on perceiving the deep particularities of another's context. Imagining oneself in another person's shoes might involve realizing overt contextual facts about their situation—emphasizing the role of listening closely in conversation and using one's accumulated experience—or employing an emotional knowledge that is sensitive to verbal and nonverbal cues. A perceptive moral imagination positions us closer to the subject, allowing us to approximate the most ideal response to the situation at hand.

As an Aristotelian philosopher, Nussbaum primarily upholds the view that moral actions that spring from certain virtues of character must involve the use of practical wisdom as well as feeling the appropriate level of emotional response to the present situation. According to Aristotle, moral virtues of character are a balance—a mean—between excess and deficiency of certain passions and actions (Aristotle 2012). For example, he says that "righteous indignation is the mean between envy and spite," and he characterizes these two latter traits as vices on the opposite ends of a spectrum (Aristotle 2012, 30). Righteous indignation, the mean emotion, is the moral virtue in this case. Accordingly, Nussbaum argues that achieving these appropriate levels of passion and expression requires some form of moral education since this knowledge of means is not an intrinsic one. This is the foundation of Nussbaum's argument for the moral value of the novel. First, by virtue of its artistic form, the novel communicates the inseparability of moral theory from the "practice aspect . . . which demands the active engagement of the interlocutor's own moral intuitions and responses" (Nussbaum 1990, 139). Nussbaum argues that, instead of presenting us with moral precepts, the novel shows morality acted out in

highly particularized situations; this is valuable because it endows readers with practical wisdom and emotional knowledge necessary for the type of virtuous action that Aristotle and Nussbaum endorse.

This is one way in which Nussbaum considers the literary form of the novel morally educational, but she also sees educational potential in the emotional "catharsis" that a novel engenders. In his *Poetics,* Aristotle repeatedly affirms that ancient Greek tragedy powerfully rouses the emotions and engenders a "catharsis" of them.[1] In this paper, catharsis is understood as "clarification," in an intellectual sense; when a reader experiences the catharsis of an emotion, it means that said emotion is made clearer to her, such that it is no longer mysterious to her and she holds an improved understanding of it (Woodruff 2008, 6). Nussbaum uses similar language to talk about the novel, writing that our emotions are in some sense guided by the skillful approach of the author to emotional themes. She argues that "affectivity is best put into play when controlled by the disciplined and essentially loving imagination of the novelist" (Kalin 1992, 137). Primarily, attentive readers are guided by the novelist's development of characters, plot, and themes to reflect on their own internal and external lives with a different perspective on the emotions, whether it be the novelist's or the characters' perspective. Moral education takes place as this process improves and clarifies the reader's own understanding of the emotions, leading to more informed processing and expression of emotions. Furthermore, according to Nussbaum and others of her opinion, this emotional clarification offered by the novelist allows readers to see their "shared vulnerability" with others in a collection of shared emotions (Woodruff 2008, 16). Ultimately, Nussbaum revives Aristotle's thought on catharsis by applying it to the more modern literary form of the novel.

Nussbaum further refines Aristotle's connection between tragedy and emotions by reference to the Jamesian idea of morality. In her essay "'Finely Aware and Richly Responsible': Literature and the Moral Imagination," Nussbaum writes that Jamesian morality stresses optimized sensitivity in everyday interactions and maintains that "obtuseness and refusal of vision are our besetting vices": that is to say, the failure to act perceptively in the moment constitutes a moral failing (Nussbaum 1990, 148, 152). This concept, of course, has an Aristotelian origin since it emphasizes not only action imbued with emotion but also the discernment of the particular context at hand (Nussbaum

---

1   Aristotle's use of the word "catharsis" has been contentiously debated since the discovery of the treatise, since the word had been employed by Hippocrates in medical contexts to denote the purgation of poison, by Pythagoreans in religious contexts to denote the cleansing of the soul, and in other contexts before Aristotle employed it in the *Poetics* (Woodruff 2008, 3). Nussbaum and other contemporary Aristotelian scholars such as Stephen Halliwell land on the translation of "catharsis" of the emotions to "clarification" of the emotions. This clarification makes something that is "intellectually obscure," like a feeling, clearer (Woodruff 2008, 6). Nussbaum writes that this clarification helps us realize the deeper and more essential aspects of who we are (Nussbaum 1992).

1990). Just as Aristotle believes that a virtuous course of action must be tailored to the particulars of a situation, in order to be practically wise, Nussbaum, too, writes that moral individuals ought to develop "the ability to discern acutely and responsively, the salient features of one's particular situation" (Nussbaum 1990, 37).

Nussbaum is clear that she does not want to dispose of more generalized ethical systems, writing that "rules and general categories still have enormous action-guiding significance in the morality of perception" (Nussbaum 1990, 37). Rather, it is that these habits of perceptiveness are a needed supplement to general rules, such as rules regarding compassion and empathy. Fixed rules will not prepare a moral agent to respond sufficiently to "new and unanticipated features" or to "the context-embeddedness of relevant features," because the ethical scene cannot be simplified (Nussbaum 1990, 38). Indeed, for Nussbaum, certain features of a literary work best diminish the chance of obtuseness, inculcating in the reader Aristotelian moral behaviors.

2. Aristotelian views on Greek tragedy and ethics

Before I proceed, I will demonstrate how Nussbaum's argument that the novel can be an ideal form of moral literature derives from Aristotelian philosophy. First, one must understand the long history of this debate about the connection between literature and morality. By expanding on this history, I will demonstrate the multivalent affinity between Nussbaum's argument for the novel as ideal moral literature and Aristotle's argument for tragic poetry as the same. Furthermore, I must show not just that the novel is an ideal genre for moral education but that any form of literature with the characteristics laid out by both Aristotle and Nussbaum is ideal, as well. This will allow me to distinguish certain forms of contemporary poetry from other forms of poetry, like ancient tragedy, and make way for me to modernize poetry's place in ethics.

Plato denounces poetry in Books III and X of *The Republic* and banishes the poets from his utopian society, despite—or perhaps, because of—poetry's prominence in the ordinary lives of Athenian citizens. The most relevant aspect of this exile for the purposes of this paper lies in Plato's rejection of the ethical value of poetry and other forms of artistic representation. Noël Carroll writes in his essay "Literature, the Emotions, and Learning" that "where those lines aroused emotions that Plato thought would potentially subvert the ability of the guardians to defend and rule the Republic, he argued that the poetry should be regulated . . .," meaning that the type of poetry that gives rise to concerns related to morality is tragedy. For Plato, hymns that faithfully represent the virtue of the gods and other types of poetry conducive to the preservation of an all-reverent and hyper-rational utopia were allowed to remain. Plato's distrust mainly stems from the tragic poets' dismantling of the supremacy of reason. In the introduction to his translation of Aristotle's *Poetics*, Richard Janko writes that in the *Ion,* one of Plato's early works, Socrates denounces tragic poetry because it "encourages the audience to

indulge their emotions, most notably pity and fear, to the detriment of their powers of reason" (Janko 1987, xi). Plato also distrusted tragedy for its engagement in *mimēsis* (i.e., "imitation," in Plato's sense of the word). As Plato explains in *Republic* X, tragic poetry engages in a thrice-removed imitation of the reality of the Forms, since the objects and events of the perceptible world are already imitations and constructions of art; this confuses and misleads us to think that representations and forms of art are actually emblematic of what exists (Janko 1987, xii). Plato argued that these representations led readers astray from reason, which he considered the ultimate good, and presented them with complex rather than straightforward systems of morality since the poets supplied unstable understandings of what makes a "good man."

Plato faced the conflation of logic and literature, as he and other philosophers in his generation thought that all communication should aim at reason and truth. Later, Aristotle was able to extricate the two, making room for the moral, civic, and educational value of emotions. Aristotle replies to Plato's *mimēsis* objection by building a philosophy around the importance of discerning particulars, rather than ascertaining knowledge of general ideas (Janko 1987, xiii). For Aristotle, poetry is a useful representation of reality and has pedagogical value, since the author selects, interprets, and emphasizes meaningful aspects of life (Janko 1987, xv; *Poetics* trans. Janko 1987, 4). Reproductions of reality allow us to indirectly experience undesirable or unattainable events. This is in line with the value Nussbaum places on vicariously undergoing a moral event and learning from it by inhabiting the particulars it presents, endowing readers with supplemental practical wisdom.

Aristotle says in the *Poetics* that readers "derive pleasure from seeing representations even of things that are painful to look at in actuality," reminding us of the possibility of catharsis, a clarification or illumination of emotions, through art (Janko 1987, xvii). Nussbaum also explores how projecting oneself into representations happens through a realistic but focused plot and the generation of fundamentally good characters. Indeed, Aristotle writes in the *Poetics* that "since [tragedy] is a representation of an action, and is enacted by people acting, these people are necessarily of a certain sort according to their character and their reasoning . . ." (*Poetics* trans. Janko 1987, 8). He expands on this by arguing that a tragedy, meant to be complex in form and content, should not show a thoroughly evil character prospering, but rather should show a good one failing due to hubris or some other fatal flaw. This failing is what Aristotle calls "terrifying and pitiable," stoking the catharsis of tragic emotions and thus offering a level of moral satisfaction from clarification (*Poetics* trans. Janko 1987, 17). Nussbaum takes from Aristotle the problematization and complication of moral representations and the value of clarifying emotions in preparing perceptive and sensitive moral agents. As I will show now, Nussbaum modernizes Aristotle's concept of catharsis by applying it to the novel.

3. Nussbaum's ideal moral literature

Nussbaum believes that the contemporary literature most suitable for the moral imagination is the novel, stating in "Finely Aware and Richly Responsible" that "[the moral imagination's] conception of moral attention and moral vision finds in novels its most appropriate articulation" (Nussbaum 1990, 148). She takes as her paradigm Henry James' *The Golden Bowl*, stating in "Flawed Crystals" that some of the features making it a paradigmatic work of both literature and moral education are "shared with other related novels and with tragic dramas" while "others are peculiarly its own, or belong to it in a particularly high degree" (Nussbaum 1990, 138). I will describe those defining literary features that characterize *The Golden Bowl* and explain how they strengthen the moral imagination and, consequently, moral behaviors of the Aristotelian kind I have outlined.

The first defining formal characteristic is lifelike elaboration, which is native to the novel genre. Realistic representations generate a full world of the novel and so promote the reader's projection into the rendered scenarios. The author of the novel usually deeply attends to particulars. Nussbaum writes that descriptions that succinctly totalize a scene actually deprive readers of "the sense of lucidity, expressive feeling, and generous lyricism" that immerses them in the piece and thus positions them as the subject or approximate to the subject of the depicted experience (Nussbaum 1990, 152). The idea of "vicarious experience" has moral import for Nussbaum, as it expands the reader's repertoire of situational knowledge and emotional understanding, supplying her with more resources to imaginatively navigate moral challenges with maximized sensitivity in real life. Lifelike elaboration in a novel also builds empathy. Nussbaum writes that "a novel, just because it is not our life, places us in a moral position favorable for perception and it shows us what it would be like to take up that position in life" (Nussbaum 1990, 162). Vast detail allows readers to visualize themselves in a situation often drastically remote from their own and familiarize themselves with it. Moreover, Nussbaum emphasizes that morally relevant exchanges—such as those between Maggie and Adam Verver, two main characters of *The Golden Bowl,* whose deep compassion for one another in the face of a dilemma threatening to separate them as father and daughter lies at the center of the novel—must be rendered with precise detail. This is because being able to immerse oneself in a character's action through the generation of tone and intricate diction *models* for us both moral behaviors and sensitivity, rather than merely relating precepts in absolute terms.

The kind of moral particularism Nussbaum discusses abides by more general inviolable rules, as I have mentioned, but goes beyond them to say that we also must appropriately address the contextual particulars of the situation before us, providing an improvisational response. This is another way the novel's world-building broadens the moral imagination: it gives us *more* than standing terms, as "a person armed only with the standing terms . . . would, even if she managed to apply them to the concrete case,

be insufficiently equipped by them to act rightly in it" (Nussbaum 1990, 156). Thus, the detailed novel provides us with highly particularized ways to approach unfamiliar, unpredictable, and even mysterious situations. Indeed, Nussbaum believes this degree of detail is a fundamental characteristic of the ideal genre of moral literature.

Nussbaum argues that, in conjunction with lifelike elaboration, a second defining formal characteristic of the novel is imperative for the strengthening of the moral imagination, and that characteristic is traditional narrative form. In "Flawed Crystals," Nussbaum highlights that the construction of complex characters and the progression of their lives in an expansive plot are crucial: "It is not only, then, the novel's capacity to explore the length and breadth of a life, but the combination of this exploratory power with the presence of a character who will count as a high case of the human response to value . . ." (Nussbaum 1990, 140). As such, the exemplary moral novel does not only offer the author's representation of a moral dilemma but also demonstrates that dilemma's gradual manifestation and resolution in the lives of round characters, which takes certain space, certain plot progression, and other structural complexities. To this end, Nussbaum implies that part of what makes *The Golden Bowl* a rare ideal work of moral literature is the characters' oscillations between virtue and human mistake, making moral error an inevitable part of moral behavior—an argument in alignment with life's messiness and one that Nussbaum suggests traditional moral philosophy would struggle to make. Indeed, she writes in her essay "Flawed Crystals" that it is only when "we study the loves and attentions of a finely responsive mind such as Maggie's, through all the contingent complexities of a tangled human life, that the force of [moral ideas] is felt" (Nussbaum 1990, 140).

Implicit here is Nussbaum's claim that part of the moral relevance of reading this kind of novel is the characters' use of their moral imaginations. As such, Nussbaum's ideal form of moral literature is at least somewhat didactic. Reading about characters employing their moral imaginations with discretion and precision, in relation to other characters in a series of actions, provides us with a template for our own parallel moral activity. We see models of moral activity and assimilate them; we are thus given ways beyond theoretical knowledge to improve our own behavior. When we improvise solutions to unfamiliar moral challenges, we can find ourselves not left entirely in the dark but able to draw on the highly perceptive conduct of those we read about. So, for Nussbaum, there must be a moral "rightness" to the characters of a novel (not a flawlessness but a consistent aim to act morally) for it to be a work of moral philosophy, which distinguishes *The Golden Bowl* from other novels. Nonetheless, together with lifelike elaboration, a traditional narrative structure works with our phenomenological experience of reading to ingrain in the reader moral habits like attentive perceptiveness, identification with characters or situations, interpretation, and so on, thereby strengthening the technical skill of the moral imagination rather than merely furnishing "knowledge that."

Thus, for Nussbaum, it is vital that we get a full sense of the characters' lived experiences in order to expand the moral imagination, and this is only possible with characters, plot, and certain length—which uniquely characterize the novel.

In addition to these two distinguishing formal features of the novel, the content of *The Golden Bowl* instills perceptiveness and sensitivity by virtue of its moral evocativeness. Specifically, this novel is concerned with emotions regarding problematized or complex representations of morality, displaying "the complexity, the indeterminacy, the sheer difficulty of moral choice (Nussbaum 1990, 142). An example would be the dilemma that Adam faces in accepting his daughter Maggie's maturation into her own sexuality and individuated life and how he overcomes it using a "subtle and high rather than simple and coarse" moral imagination (Nussbaum 1990, 152). According to Nussbaum, this scene is rendered with "sheer gleaming beauty" of language and realistic dialogue that would be devoid of meaning without knowing the history of these characters' relationship: the "immeasurable love" behind their utterances (Nussbaum 1990, 152, 154). This is a scene rich with descriptive elaboration that spins emotion out of causal connection in the novel, since readers' knowledge of the characters and plot amplifies the poignancy of the scene. These traits position the reader momentarily as the subject, feeling what the characters feel. Indeed, this very scene suggests that the "right 'basis' for action is found in the loving dialogue of the two" (Nussbaum 1990, 155). Another example of an emotionally evocative and morally pertinent scene is Maggie's struggle to arrive at a decision to either marry or stay a dutiful daughter to her loving father, her experience of "moral anguish" (Nussbaum 1990, 150). By connecting this emotionally evocative scene-building to the characters' experiences of moral dilemma, James encourages a morally enlightening catharsis in the reader, which would clarify the emotions rendered in connection to the loving treatment of others and the pursuit of the right action. Such cathartic plot-building creates a moral imagination highly tuned to the breadth and depth of the emotions, better preparing its possessor for action in particular situations that are properly imbued with and responsive to others' feelings.

III. CONTEMPORARY POETRY AS MORAL PHILOSOPHY

Given the nature of the discussion so far, we might be left with the impression that Nussbaum's ideal moral literature leaves us with a narrow category of works that fulfills this ideal. Indeed, she states that if this ideal is to be taken seriously, "it seems difficult not to conclude that we will need to turn to texts no less elaborate, no less linguistically fine-tuned, concrete, and intensely focused, no less metaphorically resourceful" than *The Golden Bowl* (Nussbaum 1990, 157). Now, I will turn to the project of expanding her paradigm by assessing the moral value of contemporary poetry, a genre that exhibits defining characteristics opposite to the two formal characteristics of the novel, which were elaboration and narrative structure. I will detail the opposite features I find in

contemporary poetry, which are abstraction and experimental structure. Then, I will consider their distinct contributions to the moral imagination. I will use Nussbaum's conception of virtuous behavior being generated by the moral imagination. As such, I will need to show how poetry clarifies emotions and contributes to sensitive behaviors.

1. Lifelike elaboration vs. abstraction

Here, I argue that abstraction is a characteristic most distinct from elaboration and is found in high degrees in contemporary poems. Achieved mainly through mainly the recurrent use of figurative devices, abstract*ness* is a general quality of a literary work that assigns content and meaning away from the concrete. It comes in degrees, of course; even *The Golden Bowl* is somewhat abstract in its use of extended metaphors (e.g., its use of the marred golden bowl to symbolize the "fracturing" of its characters' moral perfection) and its use of local metaphor, both of which work to make it artful rather than a dry representation of life and contribute to sensory immersion. But the more amplified and consistent the abstractness of the language, the more likely it is to perform or achieve a significant abstracting away of the *context* of the central situation. This is abstraction.

A highly abstract poem might still have a detectable "situation." In other words, it may depict some sort of event, revolve around a relationship, or explore a certain emotional state. However, the particular situation of the contemporary poem is often abstracted away from overt contextualizing facts and is instead constructed out of figurative connections. Where elaboration provides a view into the represented situation and guides the reader's grasp of that situation, abstraction detracts from the particulars of the situation, mystifying the reader's efforts to fathom herself *there* in the concrete world of the work. Importantly, the language that an abstract work uses is a succession of complex, enigmatic, or momentarily impenetrable metaphors and other kinds of figurative language, leaving the reader not with a sense of a world with discrete, well-defined features but with a sense of a general but immersive atmosphere. The figurativeness is consistently displayed and characterizes the work. As such, the individual instances of figurative language may create clear, particular meanings for the reader, but they may also work to generate generalizable lessons for them, as well.

Let us take a look at a contemporary poem by Louise Glück (1999) that significantly abstracts away context:

> It is not the moon, I tell you.
> It is these flowers
> lighting the yard.
>
> I hate them.
> I hate them as I hate sex,
> the man's mouth
> sealing my mouth, the man's

> paralyzing body—
>
> and the cry that always escapes,
> the low, humiliating
> premise of union—
>
> In my mind tonight
> I hear the question and pursuing answer
> fused in one sound
> that mounts and mounts and then
> is split into the old selves,
> the tired antagonisms. Do you see?
> We were made fools of.
> And the scent of mock orange
> drifts through the window.
>
> How can I rest?
> How can I be content
> when there is still
> that odor in the world?

In "Mock Orange," Glück forges an association between her disdain for mock orange flowers—which are supposedly quite aromatic and thus have a strong presence—and her hatred of sex—which she characterizes as a suffocating and humiliating act. The reader is left to grasp what connects these two ideas in the mind of the speaker. Is it the archetypal association of flowers with reproductivity, especially feminine reproductivity? Is there some promise of reproductivity that has been lost? Or is it a more fleeting association—that both are now occupying the speaker's mind, preventing her from resting at night? Additionally, the question that arises and the answer that becomes "fused in one sound" are both objectively unknowable to the reader; in fact, this is something quite difficult to visualize or conceptualize. A reader can only guess some context along the lines of the following: that this cycle of rumination in which the speaker is stuck makes her face her old habits, potentially ones that were self-defeating or self-loathing. The speaker is also distinct from her old selves, suggesting either maturation or naivety. Ultimately, one discerns that these flowers in their false promises of fruit are mocking the speaker, just as the old events are. At any rate, what readers take away from this abstract contemporary poem will be perpetually unstable and subjective, since their interpretations do not arise out of a situation in a direct context. The subject of the poem is held at a distance by these question-generating abstractions.

There is another facet of this poem (and contemporary poems in general) that contributes to the abstracting away of the situation: brevity. Whereas part of elaboration is the attempt at exhaustive description, abstraction also works by synthesizing images or

ideas into a few representative and artfully chosen words that serve to forge connections far beyond what might be expected. Brevity enables almost a limitless number of associations. It contributes to abstraction in that it excludes potentially significant context and leaves much of its construction open to the mind of the reader. Taken together with the abstractness of figurative language, brevity creates an enveloping atmosphere rather than a realistic, projectable world; it demands that the reader is involved in the work of interpretation of the formally and substantively ambiguous, so that she inhabits this atmosphere beyond the minimal words of the piece. The creation and envisioning of meaning involve more labor on the reader's part than a typical work of fiction would. The author offers up some possible world with the features that I have discussed, and the reader meets the author with her own associations in order to create a fuller picture out of the poem's sketch of a situation or even its complete lack of one. This seems similar to the universalizability that Aristotle claims poets practice in his *Poetics*. Here, the poet's work is more philosophical because it does not merely record but engenders widely applicable notions (*Poetics* trans. Janko 1987, 12). Indeed, this has to do with the vagueness of the writing and the earnest response of the reader to that vagueness, as the reader brings her own concepts to the table to fill out the picture and construct a subjective meaning.

This leads to a discussion of what abstraction and generation of a more vague, more exigently interpretable atmosphere do for the *moral* imagination, especially as juxtaposed to what the creation of an elaborate world does. When the situation is abstracted away, a couple of related phenomena are given to the reader that can ultimately lead to the expansion of what she can conceive, thus increasing her chance of behaving morally. First, the sensitive reader invests more time and labor attempting to understand the meaning of an abstract text. This translates to the reader doing more work to arrive at a text's meaning, which can potentially ingrain the suggestions of the text and the atmosphere that it engenders more deeply in her mind. In the case of "Mock Orange," the reader might gain a clearer understanding of the types of emotions that can become attached to sex for a woman living in a patriarchal society. The metaphorical concretization of heterosexual love as the mocking scent of the flowers might indeed stick with the reader for its peculiarity and the time it takes to arrive at this one possible interpretation, due to the abstraction of the "situation" away from context. Relatedly, the very quality of abstraction defers meaning indefinitely; an absolute meaning continues to elude the reader. This is congruent with Jacques Derrida's method of literary analysis, which suggests that the absolute "meaning" of a text itself does not exist or is always deferred (Derrida, 2016). Nonetheless, within a poem with a high degree of abstraction, meaning will at least be partially created through the practices of the reader and the associations that she makes—especially as the images within an abstract work provoke diverse interpretations by virtue of their consistent replacement of objective description with figurative language.

Abstraction increases the amount of work the reader must do to understand a piece, but there remains the inherent impossibility of fully understanding an abstract poem. The characteristic of abstraction in contemporary poetry informs what the reader ought to do in approaching challenges that ask the moral individual to behave sensitively. Just as the reader attempts to understand the complexity of the text's world, so should she approach the treatment of unfamiliar moral challenges with the same devotion to understanding but with a simultaneous acknowledgment of the impossibility of ever fully reaching a decisive understanding of the text. The piece of elaborate literature enables the morally sensitive reader to project herself into the represented, unfamiliar situation and emerge with tools to "empathize" or approach analogous situations in real life. On the other hand, the moral education offered by abstract pieces denies the possibility of truly empathizing with that which has never been directly experienced. Abstract works are valuable in that they acknowledge literature's ability to provide only a partial picture, thus preparing the reader for sensitive behavior within her bounds; they beget humility. The heightened but curbed perceptiveness that arises from reading abstract works of literature evidently can also supplement theories of morality.

2. Traditional narrative structure vs. experimental structure

Just as lifelike elaborateness relates to traditional narrative structure within Nussbaum's ideal moral literature of *The Golden Bowl* or similar novels, abstraction is related to experimental structure for the kind of literature in which we are interested. Experimental structure comes across in the way words are arranged visually on a page, with the use of erratic spacing or blank spaces rather than the text being arranged in a conventional block or stanza format. Experimental structure can also arise from the author's arrangement of her ideas on the page—with continuity being something more traditional and disjointedness being more experimental—or from the author's use of language by employing words in succession that do not make immediate sense together.

Indeed, it is Nussbaum who argues that form should adhere to content. In her essay "Form and Content: Philosophy and Literature," Nussbaum writes that "style itself makes its claims, expresses its own sense of what matters" and that "literary form is not separable from philosophical content, but is, itself, a part of content" (Nussbaum 1990, 3). Thus, Nussbaum asserts that most philosophical prose makes its argument implicitly for the preeminence of rationality, even while it may argue explicitly for the value of the emotions; the traditional philosophical text may, therefore, send contradictory messages to the reader. Nussbaum indicates that a well-written, emotionally provocative story following characters devoted to the good best acknowledges the unpredictable incidents of life, while also preparing the reader's moral imagination to handle similar incidents she may confront in the future. And while it is true that the type of story that Nussbaum champions is geared more toward the provision of practical wisdom than

traditional philosophical texts are, the text that is structurally speaking, more experimental, appears to better mirror the impossibility of decisive indications of what makes an action "correct." While works of fiction with their vast contextual detail purport to provide somewhat of a template of acting morally, experimental poetry may be a better fit for the subjective and disorderly nature of an individual's experience in the world. Let us now take a look at Lyn Hejinian's "[A straight rain is rare . . .]" (2012)::

> A straight rain is rare and doors have suspicions
> and I hold that names begin histories
> and that the last century was a cruel one. I am pretending
> to be a truck in Mexico. I am a woman with a long neck and a good burden
> and I waddle efficiently. Activity never sleeps and no tale of crumbling cliffs
> can be a short one. I have to shift weight favorably. Happiness
> can't be settled. I brush my left knee twice, my right once,
> my left twice again and in that way advance. The alphabet
> and the cello can represent horses but I can only pretend
> to be a dog slurping pudding. After the 55 minutes it takes to finish
> my legs tremble. All is forgiven. Yesterday is going the way of tomorrow
> indirectly and the heat of the sun is inadequate at this depth. I see
> the moon. The verbs ought and can lack infinity and somewhere
> between 1957 when the heat of the dry sun naughtily struck me
> and now when my secrets combine in the new order of cold rains
> and night winds a lot has happened. Long phrases
> are made up of short phrases that bear everything "in vain" or "all
> in fun" "for your sake" and "step by step" precisely. I too can spring.

Hejinian's poem does not experiment as much with the arrangement of words on the page as with renouncing concern about the sense a sequence of words makes. The poet appears unconcerned with the clarity of her sentences; it is a series of rather short, disconnected, and enjambed sentences, thereby creating images that are not always visualizable. For these formal reasons, and perhaps by virtue of being a Hejinian poem, this work belongs squarely within the Language poetry tradition; this tradition values the role of language itself in generating meaning and deemphasizes the role of the author in determining it. This makes it what Hejinian refers to as an "open text," meaning that it is a poem open to sundry, varied subjective readings. In her essay "The Rejection of Closure," Hejinian defines the open text as one which "invites participation, rejects the authority of the writer over the reader . . . and speaks for writing that is generative rather than directive" (Hejinian 2000). Furthermore, Hejinian states that form is a substantial part of what opens a text and animates all of its distinct details, mirroring the breadth of particulars in the world. This is akin to what G. Matthew Jenkins argues in his book *Poetic Obligation: Ethics in Experimental American Poetry After 1945*. Namely, an exper-

imental or "open" form in poetry mimics the refusal of the world to "cohere into our rational systems," the mold within which we have traditionally forced our moral principles to fit (Jenkins 2008, 216). Indeed, Jenkins suggests that such formal and lingual experiments do something similar to what I have suggested abstraction does, writing that these experiments "change the reader's role in the production of meaning" from "treating the reader as passive receptor of ethical pronouncement" to an active participant in its creation (Jenkins 2008, 223). Contemporary poetry with its experimental form encourages this type of participation to a higher degree than other forms of literature.

Similarly, the experimental form does not pretend to totalize its subject matter. The resistance to logical summation is, according to Jenkins, deeply connected to the *correct* vision we should have of the Other, that is, the individual who is not us, and the individual with whom we can enter into a dialogue when reading an experimental poem (Jenkins 2008, 225). We might even think that the Other, who appears to oppose the Self, might be the author, since, in this kind of language poetry, literal experience is not being related as much as fleeting, successive thoughts. This creates an intangible feeling rather than a contextualized situation. In the case of "[A straight rain is rare . . .]", the reader is asked both to determine the stream of consciousness and the association play at work and to accept that both of these private processes are inaccessible. For example, in the lines starting with "Activity never . . ." and ending with " . . . a dog slurping pudding," the reader might apprehend a subjective connection between a careful gait and the crumbling cliffs and between this image of a treacherous walk and the pursuit of an elusive happiness. But the image of the letters of the alphabet and the notes that a cello makes coming to represent horses, while the speaker "can only pretend to be a dog slurping pudding," is confounding and disorienting, leading the reader to reimagine the prior lines in connection to this one, which causes those connections to fall apart. This is emblematic of some persistent effort to identify with the Other, who is made abstruse and remote by virtue of the experimentality of the piece.

What comes of this is a similar effect to abstraction: the indefinite deferral of full understanding, though the textual features deferring meaning vitally demand a laborious search for it. I have already suggested that this kind of Derridean deferral of meaning results in more work for the reader to partake in the creation of meaning, resulting in the poem having a greater subjective and individual power. In the next subsection, I will further expound the tension between striving for and resisting summation, as well as the moral implications of the emotions stirred by poetry.

3. Moral takeaways: epistemic modesty and the emotions

I showed in Section 2 that Nussbaum's ideal moral novel would consistently depict characters aiming for and approximating highly sensitive behaviors and that the formal characteristics of the novel are insufficient in her framework to truly make a novel a work

of moral philosophy. She asserts that it must have a didactic component, following characters with virtue; the emotions portrayed in the novel and the emotions it evokes must be related to virtuousness. Yet I will add here that much of contemporary poetry offers highly beneficial contributions to the moral imagination solely by virtue of the formal contours I have laid out (i.e. abstraction and experimental structure).

Abstraction and experimental structure alone put the reader in confrontation with the irreducible subjectivity and nuance of individual experience, creating a perception of the Other that is fundamentally "epistemically modest." Epistemic modesty should be taken here to characterize a point of view that is always aware of its own perpetual lack of knowledge of the Other; it is the resistance to totalizing the subject that, as I have shown, abstraction and experimental structure foster. Epistemic modesty rivals the ideal of didacticism in Nussbaum's paradigmatic moral novel, as the cultivation of an epistemically modest understanding is in no way a direct or didactic method of moral education. In fact, instilling epistemic modesty is an intentionally indirect method. Nonetheless, epistemic modesty might be compatible with Nussbaum's Aristotelian and Jamesian moral framework since it lends itself to avoiding obtuseness. Epistemic modesty is a skill of the imagination; it tempers one's aggrandized confidence and reduces one's tendency to make hasty deductions when faced with complex circumstances that are unfamiliar. Whereas a novel might claim to occasion intersubjectivity or shared subjective experience between the characters and the reader, a contemporary poem, by virtue of its form, makes no claim to do this. Instead, it provides a conception of morality that prizes the skills of listening and of constantly *striving* to approximate a better understanding of another's struggle or joy, thereby leading to more considerate and compassionate treatment of the counterpart. A compassion-based ethics is not new. For example, care ethics affirms the goodness of responding to others with sensitivity and perceptiveness. Prominent care ethicists include Carol Gilligan, Nel Noddings, and Eva Feder Kittay. Notably, Kittay maintains the importance of epistemic modesty (Hamington and Rosenow 2019, 30). Such modesty promotes the Jamesian conception of morality, since it would prevent one from acting obtusely by falsely assuming a self-assured position where intersubjectivity is not possible, acknowledging the ineffability and unknowability of others' experiences. To further emphasize contemporary poetry's potential to reduce obtuseness and align with Jamesian morality, I highlight again its potential for the development of a deep and ethically valuable understanding of the subject.

In their book *Care Ethics and Poetry,* Maurice Hamington and Ce Rosenow maintain that prior knowledge is required to take appropriate care and imagine properly regarding unfamiliar moral challenges. However, the authors also suggest that poetry can achieve a kind of intersubjectivity between the reader and the subject of the poem and impart direct knowledge of the Other. Hamington and Rosenow use the example of Lucille Clifton's poem "poem to my uterus" to show that the reader's takeaway from

a poem about a hysterectomy will differ depending on the gender identity of the reader. Yet Hamington and Rosenow also suggest that a man reading this piece might—with work—approximate an understanding of what the experience of having a hysterectomy is like. They state that "differently configured beings must work harder to imagine and empathize with dissimilar bodies, and yet it can be done" (Hamington and Rosenow 2019, 51). But an epistemically modest ethics could be used to intervene here and assert that this imagining cannot truly be done, at least in totality. Furthermore, admitting that this work of the imagination is incomplete in some way is morally valuable for developing an understanding of emotional subjectivity–or the different ways in which we all experience emotions–and for fostering better care of one another. Contemporary psychological and philosophical thought on the emotions is in agreement with epistemic modesty on this very matter, as consensus grows in the direction of the inaccessibility of epistemically transformative experiences – that is, experiences about which knowledge can only be obtained by having the experience directly, such as the hysterectomy that Clifton writes about (Paul 2015, 8). This means that Nussbaum's refined conception of morality arises in contemporary poems insofar as they promote emotion *and* resistance to obtuseness to maximize the reader's sensitive moral behavior. Thus, Nussbaum's paradigm may not have to be as exclusive as it appears.

Additionally, in order to place contemporary poetry on an equal playing field with Nussbaum's paradigm of moral literature in the realm of ethics, it is necessary to show that contemporary poetry also evokes and clarifies emotions, or precipitates a catharsis of them. Again, I will argue that contemporary poetry does this by virtue of its formal attributes alone, rather than blending the didactic content of characters using their moral imaginations with emotionally evocative dialogue or scenes. Although it seems intuitive that poetry will evoke emotion, it is important that I give a reason why this is so, although, in some ways, it seems ineffable. One idea is that the abstraction of contemporary poetry encourages empathizing with the emotions of the subject presented, since poetry often generates an emotionally hued world that the subject inhabits and into which the reader cannot resist entering. Another idea is that abstraction makes room for the reader to interpose her own understandings and experiences, amplifying her emotional identification with the "situation" of the piece. This augments a reader's understanding of an emotion, since she shares an emotional experience with the subject of the piece and can assimilate that feeling into her own affective repertoire, or breadth of the emotions she understands (Carroll 2020). In other words, the reader's understanding of the depicted "feeling" is compounded and *clarified*, as she learns what causes can precipitate that emotion, what images are associated with it, and how to identify how it may manifest in the Other. The reader's emotional intelligence is enriched. Indeed, abstraction also evokes emotion in the reader due to its relation to figurative language, which is often richer, more immersive, and more thought-provoking than ordinary

uses of language. It draws the reader into the world of the piece, suspending her in the emotional tone generated.

I have shown that both contemporary poetry and Nussbaum's ideal moral novel exemplify Aristotle's ethics. To fill out this picture, let us examine "'oh antic God'" by Lucille Clifton (2004):

> oh antic God
> return to me
> my mother in her thirties
> leaned across the front porch
> the huge pillow of her breasts
> pressing against the rail
> summoning me in for bed.
>
> I am almost the dead woman's age times two.
>
> I can barely recall her song
> the scent of her hands
> though her wild hair scratches my dreams
> at night.   return to me, oh Lord of then
> and now, my mother's calling,
> her young voice humming my name.

In studying this poem, the reader is at best only able to produce a vague understanding of the situation surrounding the death of the speaker's mother. It seems that brevity rather than figurative language leads readers away from the contextual facts of this poem's situation and performs the work of abstraction here by distilling enormous pain into a few lines. The specificity of the image—the scent of a mother's hands, her figure pressed up against a railing, the "pillow of her breasts" connoting a motherly comfort—calls us as third-person spectators to picture our own mothers if we can and to recognize this as a tremendously personal and idiosyncratic memory. It is not that we cannot feel sympathy with the speaker's grief, and it is not that a man cannot sympathize with a woman getting a hysterectomy—the experience described in "poem to my uterus." Rather, the description of experience in these contemporary poems fosters an "empathy" that, through its composition, also instills an understanding of the deep subjectivity of such epistemically transformative experiences that is morally valuable.

IV. CONCLUSION

My intention has not been to depose the novel, in its generation of a world and a plot carried out by lifelike characters, as a genre of fine moral literature, nor did I intend to undermine Nussbaum's argument that *The Golden Bowl* is an example of literature that aligns with and potentially teaches Aristotelian ethics. I have not attempted to

argue that art must have some didactic component or that the artist is responsible for instilling morality in the viewer either; rather, I have aimed to show contemporary poetry's distance from the Jamesian idea that literature must have an overt moral content. Contemporary poetry also, in its own unique ways, achieves Nussbaum's ideal of moral literature in that it also fosters clarification of emotion and inculcates perceptive habits through its very textual attributes, building emotional intelligence through the clarification of sentiments and, at once, epistemic modesty about this intelligence. As such, the reader is brought into an intimate conference with the subject of the work, carrying away lessons about this subjective experience while also realizing that they are not totalizing lessons. Contemporary poetry, by virtue of its form, can teach the reader that the Other's identity is not a monolith and can prevent the reader from becoming overly confident in her knowledge of the Other. An Aristotelian ethics, which closely cherishes contextual particulars and the ethical and emotional power of poetry, is alive and well in even the most modern of literatures: contemporary poetry. This suggests that the cultivation of practical wisdom and moral knowledge is something that may be inherent in the activity of reading literature itself.

❧

BIBLIOGRAPHY

Aristotle. *Aristotle's Nicomachean Ethics*. Translated by Robert C. Bartlett and Susan D. Collins. Chicago: University of Chicago Press, 2012.

Aristotle. *Poetics I: with the Tractatus Coislinianus, reconstruction of Poetics II, and the fragments of the On Poets*. Translated by Richard Janko. Indianapolis, IN: Hackett, 1987.

Burke, Edmund. *Reflections on the Revolution in France*. Edited by L. G. Mitchell. Oxford: Oxford University Press, 2009.

Carroll, Noël. "Literature, the Emotions, and Learning." *Philosophy and Literature* 44, no. 1 (2020): 1-18. https://doi.org/10.1353/phl.2020.0000.

Clifton, Lucille. "oh antic god." In *Mercy*. Rochester, New York: BOA, 2004.

Derrida, Jacques. *Of Grammatology*. Translated by Gayatri Chakravorty Spivak. 40th Anniversary Edition. Baltimore, MD: The Johns Hopkins University Press, 2016.

Glück, Louise. "Mock Orange." In *The First Four Books of Poems*. New York: The Ecco Press, 1999.

Hamington, Maurice, and Ce Rosenow. *Care Ethics and Poetry*. Cham, Switzerland: Palgrave Pivot, 2019.

Hejinian, Lyn. "[A straight rain is rare . . .]." In *The Book of a Thousand Eyes*. Oakland, CA: Omnidawn Publishing, 2012.

Hejinian, Lyn. "The Rejection of Closure." In *The Language of Inquiry*. Oakland, CA: University of California Press, 2000.

Kalin, Jesse. "Knowing Novels: Nussbaum on Fiction and Moral Theory." *Ethics* 103, no. 1. (1992): 135–51. http://www.jstor.org/stable/2381500.

James, Henry. *The Golden Bowl*. New York City, NY: A.M. Kelley, 1971.

Jenkins, Grant Matthew. Poetic Obligation: Ethics in Experimental American Poetry after 1945. Iowa City, IA: University of Iowa Press, 2008.

Nussbaum, Martha C. "Tragedy and Self-Sufficiency: Plato and Aristotle on Fear and Pity." In *Essays on Aristotle's Poetics*. Edited by Amélie Oksenberg Rorty. Princeton, NJ: Princeton University Press, 1992.

Nussbaum, Martha C. "'Finely Aware and Richly Responsible': Moral Attention and the Moral Task of Literature." In *Love's Knowledge: Essays on Philosophy and Literature*. Oxford: Oxford University Press, 1990.

Nussbaum, Martha C. "Flawed Crystals: James's *The Golden Bowl* and Literature as Moral Philosophy." In *Love's Knowledge: Essays on Philosophy and Literature*. Oxford: Oxford University Press, 1990.

Nussbaum, Martha C. "Introduction: Form and Content, Philosophy and Literature." In *Love's Knowledge: Essays on Philosophy and Literature*. Oxford: Oxford University Press, 1990.

Paul, L. A. "What You Can't Expect When You're Expecting." *Res Philosophica* 92 (2) (2015): 1–23. http://dx.doi.org/10.11612/resphil.2015.92.2.1.

Plato. *The Republic of Plato*. Translated by Allan Bloom. 3rd ed. New York, NY: Basic Books, 2016.

Wang, Zhuo. "Our Common Sufferings: Reflections on the Ethical Dimensions of Contemporary Disaster Poetry." *Journal of Cambridge Studies* 4 (2009).

Woodruff, M. "*Katharsis* Revisited: Aristotle on the Significance of the Tragic Emotions." *Newsletter of the Society for Ancient Greek Philosophy*, vol. 8.2 (2008).

IDENTITY ACROSS POSSIBLE WORLDS: COUNTERPART THEORY AND ITS COUNTERPARTS

by MIKHAIL BELSKY

University of Cambridge

ABSTRACT ‖ *This essay defends counterpart theory within a modal realist framework. Beginning with an introduction to Lewis's genuine modal realism, and his reductive analysis of de re modal claims, I present the advantages of the counterpart relation, particularly its non-transitivity and non-symmetry. I then turn to Kripke's Humphrey objection, which challenges the aboutness of modal claims regarding individuals. I interpret this objection as being, at least in part, motivated by the lack of transworld identity in Lewis's counterpart theory. Accordingly, I consider various accounts of transworld identity and show that they are vulnerable to counterexamples which exploit the transitivity of identity. Thus, with respect to the question of 'aboutness,' counterpart theory and its rivals are at something of a stalemate. Given the unique advantages of counterpart theory, in its analysis of de re modal claims, I conclude that counterpart theory is acceptable within the scope of Lewis's modal realism.*

It is possible, but not necessary, that any sentence that you select at random from this essay is either true or false. Why? The question answers itself. Sentences *can* express statements (which are either true or false), but they don't *have* to. They can ask questions — 'Is the cat on the mat?' — or give commands: 'Sit on the mat, cat!' So, any sentence that you might pick is not *necessarily* true or false. However, I'd wager that any sentence you pick would contain words, and necessarily so. Whilst this may not be a

profound insight, there is a clear distinction between claims that are true of an object by necessity and truths that are merely contingent. A claim such as "Mikhail Belsky was born on Wednesday," is made true by an actual, real-world state of affairs; namely, my being born on a Wednesday. However, the modal claim "Mikhail Belsky could have been born on Sunday" implies a non-actual scenario that does not obtain. This scenario is plausible enough (though not for the necessitarian). Whereas the claim "I could have been born on both Wednesday and Sunday" seems contradictory; it seems to imply an impossible state of affairs. It appears that some modal claims can be true or false in virtue of non-actual referents (i.e., scenarios that do not obtain), yet we make a categorical distinction between possible and impossible states of affairs. The former limits the scope of all possible worlds. The notion of possible worlds has been adopted as a means to elucidate the semantics of modal language, as it clarifies the meanings of necessity and contingency. In essence, a necessary truth — like "2+2=4" — is true in every possible world, while a contingent truth is true in some possible worlds. This heuristic motivation engenders an intuitive view of possible worlds as abstract entities that we construct to evaluate modal claims.

David Lewis, however, proposed a radical alternative: every possible world is real. These possible worlds are *concrete* (i.e., not abstract) entities whose existence is of no lesser 'reality' than our own world (i.e., the spatiotemporal universe). By analogy, the actual world is our home and each possible world is a house. The privileged status we confer upon the "actual" world comes from our own presence in it and is not reflective of any metaphysical significance. These worlds are spatiotemporally isolated from the actual world, so no object can exist beyond the confines of its respective world. Accordingly, a *de re* modal claim — a claim about the de re modal properties of a specific object[1] — about an individual must quantify over worlds in which that individual does not exist. Lewis, however, maintains that an individual has counterparts in these possible worlds, and these counterparts serve as the referents of modal claims, thereby allowing them to be rendered true or false. Yet, Saul Kripke objected to the capacity of counterparts to preserve the semantics of modal claims about individuals:

> "If we say 'Humphrey might have won the election (if only he had done such-and-such)', we are not talking about something that might have happened to Humphrey but to someone else, a 'counterpart'" (Kripke, Naming and Necessity, p. 45 footnote 13).

In this essay, I aim to defend counterpart theory *as a feature* of Lewis's modal realism. I do not defend Lewis's concretist ontology of possible worlds. I first introduce Lewis's

theory as a reductionist account of modality that affirms the concrete existence of possible worlds. I advocate the counterpart relation on the basis that its properties (non-transitivity and non-symmetry) enable a successful reductive analysis of de re modal claims. I then show that the theory's denial of transworld identity invites objection to the aboutness of modal claims. Accordingly, I then consider alternative transworld relations that could adhere to Lewis's modal realism through qualitative or haecceitic identity conditions. I show that these identity relations are vulnerable to counterexamples that exploit the transitivity of identity. Finally, I acknowledge the concern of Saul Kripke and Alvin Plantinga that a counterpart relation feels lacking in its account of identity. Nevertheless, I conclude that counterpart theory remains acceptable as a feature of Lewis's modal realism.

Kripke's Humphrey Objection targets *counterpart theory*, which comprises Lewis's genuine modal realism. This form of realism is motivated towards a reductive account of modal claims. Occurring in ordinary language, statements which are counterfactual conditionals, such as "had x occurred, y would have followed," are implicitly meaningful. However, identifying exactly what they express proves difficult given that they do not directly refer to any worldly state of affairs, yet intuitively are also claims about the world. The semantics of possible worlds, the set of ways that the world could be, provide a useful frame of reference for modal concepts. Possible world semantics therefore accommodate the intensionality of modal concepts as they occur in ordinary language. Lewis takes possible worlds to be concrete entities that are spatiotemporally isolated from our world whilst *ontically equivalent* (i.e., just as real and concrete as our own world). The key difference between our world — the *actual* world — and these possible worlds is simply that we exist in the actual world (Lewis, 1986, pp. 2-5, 92-93). Modal claims quantify over the set of ways that a world could be, and attributions of modal properties can be justified with reference to these possible worlds.[2] Hence, the obscure notion of possibility invoked in the claim, "Humphrey *might* have won the election," is elucidated by the semantically equivalent claim, "*There exists some world in which* Humphrey won the election." Affirming all possible worlds allows the modal qualifiers "might/must" to be replaced by quantification over a domain of worlds. Thus, granting Lewis's realism allows modality to be *reducible*. Possibility and necessity are not fundamental, unanalysable features of the world. Rather, we can analyse them simply in terms of possible worlds. However, the stipulation of possible worlds as being spatiotemporally isolated from one another invites the following question: How are modal claims about individuals *in the actual world* made true by faraway, alien individuals *in possible worlds*?

Lewis's possible worlds, being concrete and distinct, lead him to reject *transworld identity*: no individual exists across possible worlds. So to say that "Napoleon might have

---

1   De re modality concerns the modal properties that an object has in virtue of itself;
      formally, De Re Modality: Some x is such that it is necessarily F.

— 22 —

been tall" is to reify a possible world containing "the existence of a tall *counterpart* of Napoleon who is not identical with Napoleon" (Mackie, 2006, p.80). This counterpart relation is one of similarity between distinct individuals across possible worlds; it is a resemblance, or multiplicity of resemblances, in both 'content and context' (Lewis, 1968, p. 114).[3] The discerning of one's counterpart is, "weighted by the importances of the various respects and by the degrees of the similarities" (ibid, p. 115). Lewis's counterpart relation is seemingly ambiguous, despite his understanding of identity being clear in that "nothing is ever identical to anything except itself"(1986, p. 192). This reflects the theory's sensitivity to context; exactly which similarities are considered important will depend on the modal claim.[4] He aligns the relation's vagueness with that of "essence and de re modality generally, [which] may be subject to pragmatic pressures, and differently resolved in different contexts"(1983, p.42). The imprecision in fully delineating the counterpart relation is an undesirable feature in itself. But given the inconstancy of what similarities are relevant between modal claims, the development of a context guided approach to determine this transworld relation is appropriate.

If the counterpart relation were an identity relation, it would be *transitive* and *symmetric*.[5] Transitivity: If p1 is identical to p2, and p2 is identical to p3, then p1 is identical to p3. Symmetry: If p1 is identical to p2, then p2 is identical to p1. The counterpart relation, in forgoing identity, lacks transitivity and symmetry. The actual Humphrey may have a counterpart (H_1), who himself has a counterpart (H_2), ... (Hn). Yet the actual Humphrey need not stand in a counterpart relation to "Hn" if "Hn" fails to adequately resemble Humphrey. Being non-symmetric, someone can have a counterpart (C_1) without himself being a counterpart of C_1. Lewis provides an example of a possible world containing a person (C) who uniquely resembles a blend of two brothers (A, B) in the actual world. If (C) resembles (B) more than (C) resembles (A), then (B) is the counterpart of (C), whilst (C) is the counterpart of both (A) and (B) (1968, p. 116). Free from the constraints of strict identity (transitivity and symmetry), counterpart theory can accommodate a wider range of modal claims. It can account for modal claims that don't suggest the possible individual is the same as the actual individual being

referred to. For example, the saying, "If I were you ..." rarely carries a connotation of the speaker being physically or essentially transformed but rather invokes a narrow similarity between the situation of the speaker's counterpart and that of her actual interlocutor. Nevertheless, the non-identity of the counterpart relata does conflict with the assumed identity of individuals in modal claims.

Kripke objected to the relevance of a counterfactual about Humphrey's presidency given that "Humphrey could not care less whether someone else, no matter how much resembling him, would have been victorious in another possible world" (1980, p. 45). This objection can be understood as primarily targeting the *aboutness* of Lewis's counterpart relation. The question at the heart of Kripke's objection is this: Do modal claims that refer to an individual x — when analysed in terms of some possible individual y — remain meaningful claims *about* that very same individual x? Kripke finds the counterpart relation to be "more bizarre than the usual notions of transworld identification that it replaces," because Lewis's analysis of modal claims about individuals fails to properly involve the relevant individual himself (ibid.). Therefore, on Lewis's counterpart theory, modal claims are entirely irrelevant to the individuals that they implicitly refer to. Kripke's objection represents the notion that "If it is true that X is possibly not G, then our modal intuitions tell us that this statement is about that very same X and not about some counterpart of X" (Borge, p. 271). Therefore, Lewis's theory of counterpart relations appears to be incompatible with the intuitive assumption that modal statements about a person concern only that particular person.

Kripke's objection is motivated by the loss of identity in Lewis's analysis. Modal claims about an individual are analysed in terms of possible worlds *without* that individual existing across possible worlds. If we reject counterpart theory on these grounds, then we should endorse a *transworld identity* of some kind. Transworld identity refers to the existence of an individual across possible worlds. Applied to the Humphrey case, the claim, "Humphrey might have won the election if ...," would similarly quantify over the possible worlds in which Humphrey won the election. Yet the Humphries in these worlds would be identical to the world-bound Humphrey. As such, being identical does not imply a qualitative indiscernibility. Rather, these Humphries would comprise a singular identity — a "transworld mereological sum" — a composite of each Humphrey ($H1$ - Hn) extended across all possible worlds (1986, p. 197). Following from Lewis's concretism — whereby each individual exists in space and time — no accidental (i.e., non-essential) properties could comprise this identity: one's name, physical features, and personal history are all contingent facts about oneself and therefore cannot comprise a coherent identity without an apparent contradiction. Ted Sider exemplifies the contradiction that this would lead to:

> My right hand has five fingers, but it might have had six. Given transworld iden-

---

3   As amended in Lewis, 1971, p. 203

4   For example: "in one context we may favour the counterpart relation that (e.g.) weighs being two-legged more heavily than being rational, but in another context we may reverse the relative weighting of these respects of similarity, and so on" (Beebee, 2015, p.229).

5   A relation R is transitive if: xRy and yRz ⇒ xRz.
     For instance, let R be the relation '—is bigger than—'
     If x is bigger than y, and y is bigger than z, then x is bigger than z.

     A relation R is symmetric if: xRy ⇒ yRx
     For instance, let R be the relation '—weighs the same as—'
     If x weighs the same as y, then y weighs the same as x.

tity, my hand itself must somehow have five fingers with respect to the actual world while having six fingers with respect to some other world. (Sider, 2006, p. 3)

The key point here is that the accidental properties of an individual can vary across possible worlds. If an individual's identity is informed by their accidental properties, then their transworld mereological sum (identity across worlds) could exhibit contradictions. Insofar as a transworld identity cannot comprise such contradictions, to include any accidental properties in a criterion of identity would be incompatible with Lewis's realism.

Any criterion of transworld identity that is derived from an individual's essence is also problematic. An individual essence is a qualitative property, or set of qualitative properties, such that "each property in the collection is essential to the object under consideration and it is furthermore not possible for a numerically distinct object to satisfy every member of the collection" (Koslicki, p.114). Consider Chisholm's paradox, wherein one begins with two qualitatively distinct, actual individuals: Noah and Adam (Chisholm, 1979, p.83). Their non-essential properties are incrementally swapped across possible worlds, holding their individual essences constant, until Noah and Adam's non-essential properties have been entirely transposed. After this transition, the new Noah would possess the sum of the original Adam's accidental properties (physical appearance, history, behaviour etc), and vice versa. The new Noah would be qualitatively indistinguishable from the original Adam, as would the new Adam from the original Noah. Yet the essential properties of Adam and Noah remain unchanged. The transitivity of identity, along with the contingency of Adam and Noah's recognizable characteristics, would permit this interchange. However, the non-transitivity of Lewis's counterpart relation would not be forced into this absurdity.[6] The incremental deviation from Noah's extrinsic properties would gradually diminish the degree of resemblance between Noah and his counterpart. At first, the qualitative distinctions (in accidental properties) may be slight enough to preserve the counterpart relation. But a possible individual bearing no resemblance to the actual Noah could not serve as Noah's counterpart. These cases show why establishing a transworld identity with property-based criteria is problematic. Concurrent with Lewis's concretism, a transworld identity cannot supervene on (i.e., depend for its existence on) qualitative facts nor be determined indexically by the individual in the 'actual' world. Alternatively, a transworld identity relation may be described non-qualitatively as a transcendent haecceity.

In comparison with Lewis's counterpart theory, the stipulation of a person's haecceity is an unconvincing response to the problem of transworld identification. Haecceitism posits that entities have a unique, irreducible property that makes them distinct from

---

6   That is, the absurdity of Noah (N) having some counterpart (Nn) who is qualitatively indistinguishable from Adam (A), yet not a counterpart of Adam.

all other entities, often referred to as "haecceities" or "thisness. In short, a haecceity is the property of being a distinct, particular individual. For Salmon, "the haecceity of an individual x is the property of being identical with x, i.e. the property of being that very individual" (1996, p. 204). Haecceitism about individuals legitimizes the notion "that individuals can be extended in logical space (i.e., through possible worlds) ... and that a common 'thisness' may underlie extreme dissimilarity or distinct thisnesses may underlie great resemblance" (Kaplan, 1975, p. 722). A proponent of haecceitism holds the view that a possible world could be qualitatively indistinguishable from the actual world yet still differ in important ways (Skow, 2008, p. 98). From this perspective, an individual's identity does not supervene on the set of qualitative descriptors that apply to them. Whereas the counterpart theory avoids the problem of transworld identification by denying that individuals occupying distinct worlds can be identical, a haecceitic approach does not recognize a problem to begin with. If an individual is taken to have a primitive property of being identical to herself, and that property cannot be reduced to qualitative description, then the "criteria for trans-world identity are to be replaced by stipulation" (Salmon, 1996, p. 210). The statement "Humphrey might have won the election if x," would therefore quantify over the possible worlds in which Humphrey exists. In this sense, accepting haecceitism does allow for a reductive interpretation of modal claims and hence a constant reference to Humphrey himself. Rather than being conditional upon vague identity criteria, transworld identity simply holds in virtue of Humphrey's haecceity; thus a domain of possible Humphries is established plainly by stipulating their haecceity. This stipulation may be merely semantic, making haecceities implicit in all modal claims about an individual; but it may also be ontological, whereby a haecceity is posited as a truly primitive property that determines identity across possible worlds.

If an individual's haecceity is primitive and determinant of his identity, then his identity across possible worlds could not depend on any qualitative properties. Haecceitism is incompatible with the *identity of indiscernibles* because it rejects the notion that qualitatively indistinguishable objects are identical (Legenhausen, 1989, p. 626). If a haecceity is both a non-qualitative property and sufficient for an individual's identity, then by extension, no qualitative properties would be sufficient for the non-identity of that individual. The following example will attempt to show that haecceitism, both as a semantic framework for stipulation and as an ontological commitment, leads to a counterintuitive analysis of modal claims. These are claims that violate aboutness (i.e., an intuitive understanding of a subject's identity) more so than what we saw in the Humphrey example.

Under haecceitism, the counterfactual conditional, "Had Donald Trump been born in Scranton, he would have won the 2020 election," quantifies over the worlds in which Trump exists and was born in Scranton, and is a meaningful claim about Trump. Moreover, the counterfactual conditional, "Had Donald Trump been x (x being the exhaustive set of qualitative facts about Joe Biden), he would have won the 2020 election," is mean-

ingful if a haecceity is considered a primitive property, but meaningless if a haecceity is merely stipulated in order to set out a domain of possible worlds. In other words, if Trump's haecceity were merely a stipulative term, then the later counterfactual conditional is rendered meaningless since the conditional fails to pick out a possible individual. In this case, the domain of possible worlds in which a given individual exists is limited by some degree of qualitative resemblance (and not demarcated purely by the stipulation of their haecceity), and consequently, the problem of transitivity reemerges: Presumably, a moderate haecceitist would accept the first counterfactual's assumption (Trump + 1(Biden Quality)) as being a claim about the world-bound Trump whose identity holds across these two worlds. By extension, then, she would also accept a claim about (Trump + 1(Biden Quality)) where the counterfactual assumption picks out (Trump + 2(Biden Quality)), and by the transitivity of identity, this would eventually extend to the identity of (Trump + (the entire set of Biden's qualities)), contradicting the initial rejection of the possible worlds in which Trump is qualitatively identical to Biden. Therefore, even if it is not the case that haecceities are metaphysically asserted, it seems they must be stipulated as being unrestricted by qualitative facts to avoid the problem of transitivity.

If haecceity is taken as a primitive property, then the assumption of Trump's identity in any counterfactual scenario, regardless of his qualitative properties, would be admissible. So, in saying, "Had Trump been x (where x represents the entire set of qualitative facts about Biden), he would have won the 2020 election," we are still talking about the world-bound Trump. Assuming that the set of contingent, qualitative facts about Biden include his name, his physical attributes, and his personal history, in ordinary language this would be akin to saying, "If Trump were Biden, he would have won the election". However, if a purely haecceitic transworld identity were adopted, then the individual picked out in that counterfactual assumption would still be Trump. Comparatively by Lewis's counterpart theory, in saying, "Had Trump not done y, he would have won the 2020 election," we are making a claim irrelevant to the actual Trump merely because the "Trump" referenced in this counterfactual is in fact a qualitatively resemblant, non-identical counterpart. But in comparing the individual picked out by the counterpart theory's antecedent (a qualitatively indistinguishable Trump counterpart) to the individual picked out by haecceitism (Biden with Trump's "thisness"), it appears that the counterpart relation produces a claim that is far more relevant to the actual Trump. Thus, as an alternative to the counterpart relation, a haecceitic account of transworld identity is liable to violate the aboutness of modal claims.

Theories of transworld identity, just like counterpart theory, seek to explain how modal claims can quantify over possible individuals (i.e., over some and not others). Yet the aforementioned problems, with construing identity in terms of non-essential properties, seem to require the reliance upon a primitive identity relation. This identity relation could conceivably hold between two individuals (actual & possible) that bear no

resemblance to each other. A modal claim would thereby be satisfied (made true) by a possible individual who would appear entirely alien to the actual individual in question. These difficulties with transworld identity do not remedy those engendered by the counterpart relation. Counterpart theory does have counterintuitive implications of its own. For instance, in Kripke's objection, it is indeed implicit that in speaking counterfactually about Humphrey we mean to talk about the actual Humphrey, not some distant counterpart. Additionally, Alvin Plantinga alleges a paradoxical implication of the counterpart relation and the law of identity. The proposition "Everything is identical with itself" predicates every object with self-identity. Lewis furthermore holds an object's essential properties to be those which are possessed by each of its counterparts (Hall, 2010, p. 4). Socrates therefore has the property of "being identical with himself," and "being identical with Socrates." Yet, the non-identity of the counterpart relation would allow Socrates to have counterparts that qualitatively resemble him but differ slightly. These counterparts would have the essential property of self-identity but would not be identical with Socrates, and therefore lack the property of "being identical with Socrates" (Plantinga, 2003, p. 86). Therefore, Socrates would have the essential property of being self-identical but not the essential property of being identical with Socrates, which appears to be a strange consequence. Plantinga's chief concern with the counterpart relation is its suggestion that one lacks the essential property of being identical with oneself. Likewise, by Kripke's objection, the counterpart theory intuitively seems insufficient since it does not fully account for an individual's identity when converting modal claims about that individual's identity into quantified ones. Sider identifies these counterintuitive judgements as paradoxes of analysis, which reflect the susceptibility of our judgements to the particular characterisation of a counterpart (Sider, 2006, p. 2). In particular, the Humphrey objection overestimates the degree to which introducing counterparts affects the meaning of de re modal claims. Counterpart theory does not simply convert claims *about* actual individuals into claims *about* counterparts. Under a Lewisian analysis, the claim "Humphrey might have won if he had done x (*x standing in for a counterfactual course of action Humphrey could have taken*)" is in some respect talking about Humphrey's counterpart. But more importantly, the claim is also talking about the relation between two distinct states of affairs: x and Humphrey's intended aim, namely, winning the presidency. The Humphrey counterpart is functionally important, but the claim still bears upon the real Humphrey. The aboutness concern regarding counterpart theory seems to partially hinge on an overestimation of the role that the counterpart plays in a modal claim, which, considering the untenability of the counterexamples entailed by transitive identity relations, should not warrant the theory's rejection as a feature of Lewis's modal realism.

In this essay, I have argued that counterpart theory is acceptable *as a feature of Lewis's modal realism*. I first introduced Lewis's genuine modal realism and its semantically reductionist approach to modal concepts. I then applied modal claims about individuals

and presented the advantages of the counterpart relation, namely its non-transitivity and non-symmetry. I interpreted Kripke's Humphrey objection as disputing the aboutness of modal claims regarding individuals, before explaining why the counterpart relation may give cause for concern. Within a modal realist framework, I assumed that the predominant alternative to the counterpart relation was a theory that could answer the demand for transworld identity. I argued that a property-based criteria, and the stipulation of haecceities, both struggle to conserve the aboutness of modal claims. This was supported by counterexamples to the transitivity of transworld identity that are, at the very least, no less counterintuitive than those employed by Kripke and Plantinga against the counterpart relation. Granting its ontology, Lewis's reduction of de re modality is not compromised by its preclusion of transworld identity. Opposition to counterpart theory is better directed at the source, the indulgent ontology that it stems from, rather than alleging a semantic failure that would be remedied by the adoption of transworld identity.

❧

BIBLIOGRAPHY

Beebee, H., & Macbride, F. (2015). *De Re Modality, Essentialism, and Lewis's Humeanism.* In B. Loewer, & J. Schaffer (Eds.), "A Companion to David Lewis" (pp. 220-236). John Wiley & Sons Ltd.

Borge, S (2008). *Counterpart Theory and the Argument from Modal Concerns.* Theoria. Vol.72, Issue.4, pp.269-285.

Chisholm, R. (1979). *Identity through Possible Worlds: Some Questions.* Noûs, 1: 1–8.

Hall, N. (2010). David Lewis's Metaphysics. Stanford Encyclopedia of Philosophy, Fall 2010.

Kaplan, D. (1975). *How to Russell a Frege-Church.* Journal of Philosophy. Vol. 72, pp.716-729.

Koslicki, K. (2020). *Essence and Identity.* In Mircea Dumitru (ed.), "Metaphysics, Meaning and Modality: Themes from Kit Fine". Oxford University Press. pp.113-140.

Kripke, S. (1980). *Naming And Necessity.* Harvard University Press.

Legenhausen, G. (1989). *Moderate Anti-Haecceitism.* Philosophy and Phenomenological Research. Vol. 49, No. 4, pp. 625-642.

Lewis, D. (1968). *Counterpart Theory and Quantified Modal Logic.* The Journal of Philosophy. Vol. 65, No. 5, pp. 113-126.

Lewis, D. (1971). *Counterparts Of Persons And Their Bodies.* Journal of Philosophy, Vol. 68, No. 7, pp. 203-211.

Lewis, D. (1983). *Philosophical Papers: Volume I.* Oxford University Press.

Lewis, D. (1986). *On The Plurality Of Worlds.* Oxford: Blackwell.

Mackie, P. (2006). *How Things Might Have Been: Individuals, Kinds, and Essential Properties.* Oxford Clarendon Press.

Plantinga, A. (2003). *Essays in the Metaphysics of Modality.* Oxford University Press.

Salmon, N. (1996). *Trans-World Identification and Stipulation.* Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition. Vol. 84, No. 2, pp. 203- 223.

Sider, T. (2006). *Beyond the Humphrey Objection.* philpapers.org.

Skow, B. (2008). *Haecceitism, Anti-Haecceitism and Possible Worlds.* The Philosophical Quarterly. Vol. 58, No. 230. pp.98-107.

[ACTUALLY] POSSIBLE

by ARNOLD DINH

University of Otago

ABSTRACT ‖ *Bricker presents us with the Skepticism of Actuality.[1] The Skepticism of Actuality argues that specificities of our reality are more nuanced than the superficialities our everyday language can articulate. Simply, the argument asks how can any given individual in any given world discern their world from any other possible, nonactual world - how do you know your world is actual and more than a mere possibility? Bricker's argument confronts all doctrines that use sets and universals[2]. Most, if not all doctrines, accept either and sometimes both. The two doctrines used to answer this question are Actualism[3] and Lewisian Modal Realism (LMR)[4] – the former grounded in actual entities while the latter in possible ones. Using these two doctrines, the answer to the question arrives at the tail end of a trilemma. The trilemma presents itself because of a key component; the actual operator. The actual operator is how the term 'actual' operates differently depending on the doctrine. Once the difference is clear, we argue that LMR has an advantage over Actualism. More specifically, a component of LMR; counterfactuals, are more direct than a component of Actualism; essences. The idea that counterfactuals are*

---

1   Bricker 2006.
2   Bricker 2006. p53-56.
3   Menzel 2021.
4   Lewis 1986.

more advantageous appears after we discuss how the ontology of LMR subsumes Actualism's. The chief focus of this thesis is that Actualism and LMR share an ontological relationship where the actual world of Actualism is indistinguishable from a possible, 'actual' world of LMR; therefore, Actualism and LMR are not mutually exclusive. Succinctly, this thesis supports that we actually exist amongst a multitude of possible concrete worlds.

INTRODUCTION

This discussion of multiple worlds started with Leibniz (Discourse on Metaphysics).[5] Like most discussions of multiple worlds, this one questions how those worlds exist. Some philosophies use sets; others, universals, and there are those that use both. Stemming from Leibniz's idea of multiple worlds, Bricker sees sets and universals as problematic. How does one discern if they're merely a member of a set or a form of a universal? The question naturally extends to the nature of the world of the one asking the question. If you're anything like me, you might wonder, how have we travelled into space and touched the moon, created advanced technologies that fit in our pockets, manufactured several modes of transportation to traverse the earth, water, and space, but remain uncertain of the nature of our reality?

You may be cleverer than this writer - I've only managed to read a few books. You might think, "What is there to be uncertain about, of course we're actual and so is our world." Then I would have to ask, "Actual in what sense?". Our discourse starts with §I and how the actual operator is used differently between Actualism and Lewisian Modal Realism (LMR). Actualism, is the idea that argues, "Everything there is exists, or is actual" - that is, nothing that exists is nonactual.[6] When using the term actual from the Actualist's perspective nothing unusual emerges. Actual is such a common word we use it without a second thought. On the other hand, LMR is an idea of possible worlds. Briefly stated, "There are so many other worlds, in fact, that absolutely every way that a world could possibly be is a way that some world is...".[7] Using 'actual' from this perspective is unorthodox and reasonable. This version of the term may require some explanation, but it's easy to understand. Once we establish the difference between the two methods of using the term, we find ourselves in a trilemma. Without revealing too much, the three options simply stated, are: 1) Accept Actualism reject LMR. 2) Accept LMR reject Actualism. 3) Accept both. §I details how we arrived at this trilemma, why

---

5   Rodriguez-Pereyra 2020.
6   Menzel 2021.
7   Lewis 1986. p2.

it's a trilemma, the details of the options, and why 3) is the best.

An old television commercial shows a scene of two passionate political parties [at the brink of a civil war] arguing over who's idea is better. From nowhere, a young girl innocently walks in between the dispute and casually offers the solution "¿Por qué no, las dos?" (-Why not both?-). Immediately, the two parties drop their weapons and celebrate the wisdom this random girl presented to them. §II argues why option 3) is the best. Option 3) allows us to accept both ideologies with the provision that the ontology of Actualism is subsumed by the LMR. Ontology is the study of what exists. Both philosophies include possible worlds. Possible worlds for the Actualist exist in a less direct way than they do for the LMR-ist. Nonetheless, they exist in both doctrines. The main distinction is they exist abstractly for the Actualist and concretely for the LMR-ist. The distinction between abstracta and concreta is a complex discussion itself, but the 3rd option allows us to bypass the conversation altogether. Bypassing the distinction of concreta and abstracta allows us to focus on how the actual world of the Actualist is one of many 'actual' worlds of the LMR-ist and illustrates how the Actualist's ontology is subsumed by the LMR ist's. This subsumption allows us to keep all the desirable parts of both philosophies and scrap the undesirable ones – "¿Por qué no, las dos?".

The nature of multiple worlds conversations can be... ...out of this world. As eccentric as the topic may be, we can still have a productive conversation within a scholastic context. Outside the philosophy room, multiple world conversations have less structure and consistency. Ideas, rules, and facts are close to, if not completely, absent – this leads to unproductive conclusions. Inside the philosophy room we have peer-reviewed doctrines that provide consistent ideas, rules, and modal claims that keep our facts in check – this leads to productive conclusions. Modal claims are statements of fact.[8] No matter how unique the claim, the constituents of the claim account for its verity. Actualism and LMR, have their own methods to address modal claims. For the actualist, essences are used. Alternatively, the LMR-ist uses counterfactuals. §III discusses how each doctrine answers modal claims, what these methodologies are, how the counterfactuals have an advantage over essences and potential problems with LMR.

I. PRELIMINARIES: POSSIBLY PROBLEMATIC

Before the trilemma is discussed some groundwork needs to be established. Three conversations that need to be understood are 1) Actualism vs Possibilism debate, which we refer to Timothy Willams.[9] 2) How LMR is a particular kind of Possibilism, as provided by Menzel,[10] and 3) how the term actual operates between Actualism and

LMR. According to Timothy Williamson the debate between Actualism and Possibilsm is "silly" because of the poor clarity and lack of distinction.[11] However, the reason we refer to his expertise is because his research provides a relevant piece of information that supports the focus of this thesis; non-mutual exclusivity of Actualism and Possibilism, which supports the subsumption of Actualism by LMR. Briefly, the argument goes like this: the necessitist can obtain a meaning with 'cash value' from the contingentist, but not vice versa.[12] In his work Necessitism represents Possibilism, or more relevantly; LMR, and Contingentism, represents Actualism. Williamson argues that the 'cash value', or the meaning, a necessitists can extract from a contingentists is asymmetrical. As a necessitist, you can find truth in what a contingentist claims; however, the truth for the contingentist inconsistently reciprocates. When we superimpose the conclusion from Williamson's work to this thesis, it suggests the two doctrines are not mutually exclusive. This may seem counterintuitive given the nature of Actualism and Possibilsm. The same asymmetry goes for Actualism and LMR; and for this thesis, the overlap is LMR subsuming Actualism. Menzel argues between Actualism and *Classical* Possibilism. The classical possibilist distinguishes between being and existence; 'existence' being synonymous with 'actual'. Lewis on the other hand, rejects this distinction. Even though this rejection disqualifies Lewis from being a classical possibilist, the distinction he makes between existence and actuality qualifies him as a possibilist. Although the LMR understands 'being' and 'actuality' to be synonymous, there is a distinction between 'existence' and 'actuality'.[13] Arguments later in this thesis will be drawn from Menzel's work. So even though the classical possibilist that Menzel argues is incompatible with Actualism is inapplicable to LMR. The catch; the term 'actual', carries a different connotation when used by the LMR-ist.

'Actual' in the LMR-ist sense is different than actual in the actualist sense. For the actualist, actual is to exist – you, me, the material from which you read this thesis, etc., all exist, meaning we are actual. For the LMR-ist, 'actual' is an indexical. An indexical is a sign indicated by an element within the context. Indexicals are context dependent. Terms like there, here, nearby, later, etc., depending on the context in which an indexical is used it may convey a completely different meaning. Whether referring to possibilities, actualities, being and/or existence, etc., both doctrines incorporate a way of distinguishing the possible from the actual. The different usages of actual confront us with a trilemma because we must decide how to distinguish the actual from the possible. Even though there are two different ways of using actual, a third option is that we can forgo the term altogether.

---

8   Garson 2021.

9   Williamson 2010.

10   Supplement to *Actualism* Menzel 2021

11   Williamson 2010. p8.

12   Williamson 2010.

13   Supplement to *Actualism* Menzel 2021.

1. Skepticism of Actuality

We find ourselves in a trilemma because as discussed in Bricker 2006, the problem of 'Skepticism About Actuality', confronts all doctrines that accept sets, universals, and/or both.[14] Although Bricker rejects LMR, he accepts a similar doctrine of possible worlds; Leibnizian Realism. Bricker calls Leibniz's *Discourse on Metaphysics* Leibnizian Realism. *On the Plurality of Worlds*, Lewis considers Leibnizian Realism under the alias of Pictorial Ersatzism. The conclusion was Pictorial Ersatzism is like LMR; perhaps even better, except it was unable to account for the contingency of actuality.[15] Herein lies the problem of the Skepticism about Actuality that we're concerned with: for the individuals of these possible worlds, how can they/we determine the world existing is the actual one? Indirectly it follows, how do doctrines that accept sets and universals – which Actualism and LMR do – determine if individuals are specific members of a set or a particular universal, or more provocatively, how do we know we are actual? We can sympathize with Bricker's skepticism, because when we consider sets and universal, we must also consider more than just the metaphysical aspects of his argument. Sets and universals raise daunting epistemological concerns too. However, before we can address those questions, we need to decide which of the three options provides the most appropriate lens to make those judgements.

2. An "Actual" Trilemma

Both doctrines use actual to discern actual from possible. As an actualist, actual entities exist in the actual world and all possible entities either obtain or are unobtainable. Possible entities that obtain are actual or will actualize and unobtained possibilities could have obtained; however, remain unobtainable, leaving their essence intact but never actualizing. As a LMR-ist, all possible entities are 'actual'. To avoid confusion, the LMR-ist is not saying possible means actual. The first two options of the trilemma are: actual as a rhetorical device, or as an indexical. Firstly, 'actual' as a rhetorical device usually prevents misapprehension.[16] For example: If someone responds with an incorrect answer; the rhetorical device would suggest, "Actually, the answer is *insert correct answer*." The initial speaker gives an answer, and the interlocutor corrects the answer. Alternatively, 'actual' in this way also acts as a tool to back-reference from a nonactual possible world to the actual.[17] I.e., Clinton could've won the 2016 US Presidential Election, but Trump actually won. Meaning, a nonactual possible world envisioned Clinton as POTUS, but in actuality, Trump was elected. In either case, both instances are used by the actualist to ground the actual with the veridical.

---

14   Bricker 2006. p53-56.
15   Bricker 2006. p42-43.
16   Humberstone 2011. p931.
17   Humberstone 2011. p931.

The second kind of 'actual' is the indexical, otherwise known as the deictic. Actual as an indexical is dependent on the context. We use indexicals in our daily speech: I, you, there, here, later, soon, etc.. Depending on the context, those words mean different things. When I say, "I", I mean, me. When you say "I", you mean, you. If I say, "I'm 'here'"; I mean, I'm at the place I'm at. When you say "I'm 'here'"; you mean, you're at the place you're at. Although some indexicals could mean the same things (here, there, etc.) just as often, they mean different things. For the LMR-ist, the indexical form of 'actual' is the common usage. Earlier it was mentioned that the LMR-ist does not attempt to equate actual and possible. As a LMR-ist, all possibilities are 'actual'. However, because actual is used as an indexical, what may be 'actual' is context dependent. Suppose there are multiple worlds. At World A, I wrote this thesis, and you are reading it. At World B, you wrote this thesis, and I am reading it. For the denizens of World A, their world is actual, and World B is possible. For the denizens of World B, their world is actual, and World A is possible. In either case, depending on context and perspective, both possibilities are actual.

So, our first two options of this trilemma are actual as a rhetorical device or as an indexical. The rhetorical usage being more common to the actualist, and the indexical which is more common to the LMR-ist.

3. Foregoing Actual

The last option of the trilemma is to omit actual altogether. Yalcin 2015, argues 'actual' as an indexical is nonexistent. Crossley and Humberstone (C&H) 1977, argue that actual as a rhetorical device is redundant. Yalcin claims that the indexical form of 'actual' is simply a tool invented to be useful in a variety of philosophical contexts and that it's a special case of the rhetorical usage.[18] Although it may be useful, Yalcin argues 'actual' as an indexical is 'unnecessary' and 'overgenerates'.[19] Additionally, C&H claim actual in the rhetorical sense is either directly or indirectly redundant.[20]

Yalcin argues 2 reasons why 'actual' as an indexical is nonexistent: 1) truth conditions exist regardless of the term actual and 2) normal propositions would be burdened with extra and sometimes superfluous meaning.[21] Actualists can sympathize with Yalcin. Regarding the first reason, for truth conditions in the actual world, prefacing everything with 'actual' would be superfluous. "I'm going to the store", is just as sufficient as "The 'actual' me is 'actually' going to the 'actual' store." The truth conditions for "I'm going to the store", are I; someone, and if they are traveling to an establishment that is a store.

---

18   Yalcin 2015. p185.
19   Yalcin 2015. p185.
20   Crossley & Humberstone 1977. p16-17.
21   Yalcin 2015. p185.

Prefacing each constituent of the sentence with 'actual' is unnecessary within the actual world. In the actual world, there is an I; me, and I'm travelling to an establishment called the store. 'Actual' fails to add any substance or property to the sentence or any of the constituents, because in the actual world, by definition, everything is actual.

Regarding the second argument, 'actual' as an indexical is a tool invented for 'modal backreferencing' from a nonactual possible world to the actual world.[22] So, when using 'actual' to modify another word in a sentence; "I'm going to the 'actual' store." the sentence becomes more burdensome than it needs to be. "I'm going to the 'actual' store.", unnecessarily implies that there is another store, and somehow, it's either nonactual, or a possible establishment other than a store. More complicated-ly, and equally as accurate in description; without presupposing Actualism, there could exist a store in another 'actual' world, and the person is referring to that store. That store in another world could be a nonactual possibility, and that nonactual, possible store is different to the store that the utterer would shop at in a different possible but, 'actual' world - we begin to see how this "overgenerates". If this explanation became unnecessarily difficult to follow that's because that's what Yalcin is arguing. For an actualist, there is only one world, and it's the actual world. By adding 'actual' to describe the store complicates by over generating possibilities that are unobtainable or are simply nonexistent. "I'm going to the 'actual' store." is equal to "I'm going to the store.". In both examples, there is a person going to an establishment called the store - no possible worlds required, no counterparts included, no extras needed. Whether we agree with Yalcin's arguments or not, we have reason to believe 'actual' as an indexical either fails to add any substance or becomes overcomplicated - both of which are sufficient reasons to reject the indexical form.

C&H argue that actual is either a directly or indirectly redundant.[23] Direct redundance is 'deleting all occurrences of the operator' and indirect redundance is 'any formula of the language in which the operator occurs, there is some provably equivalent formula in which it does not'.[24] In other words, to observe direct redundance, we would remove the term altogether. In the actual world, "my phone" is no different than "my actual phone". When comparing the former to the latter, actual fails to add any more properties than what it inherently possesses.[25] To observe indirect redundance, every

---

22   Humberstone 2011, Yalcin 2015.
23   Crossley and Humberstone 1977. p16-17.
24   Crossley and Humberstone 1977. p13-15.
25   If this example sounds like an echo of the first argument from Yalcin, that's because Yalcin argues from Crossley and Humberstone that the indexical is a special case of the rhetorical use. So even though Yalcin was arguing to omit the indexical form, because he believes the indexical is as special case of the rhetorical, we know that there is reason to accept the rhetorical form. 26 Bricker 2006. p56.

instance where the actual operator exists, a negation of that instance also exists, "my actual phone" complements "my not actual phone" and then we basically revisit the same previous conundrum, as previously outlined, where actual fails to add any property - "My phone" complements "my not phone" sufficiently and actual is unnecessary. Even though there may be a preference of indirect over direct validity or vice versa, ultimately; in both cases, the term is redundant. Whether we agree or disagree with C&H, there is a reason to reject actual as a rhetorical device: simply, because it's redundant.

In summary, Bricker argues, "If there were *a priori* grounds for holding that everything is actual, as actualists believe, then skepticism could be defeated; but there aren't. If actuality were relative rather than absolute, as Lewis believes, then knowing that one is 'actual' at one's own world is all there is to know; but it isn't.".[26] These arguments raise concerns for actuality. Recap, Yalcin argues the indexical overgenerates, complicates, and is unneccesary - we have reason to reject the indexical case. At best, C&H argue actual is some sort of redundancy, but redundant nonetheless - we have reason to reject the rhetorical usage. These arguments support Bricker's skepticism. If both doctrines are unable to address the skepticism of actuality, then we may just have to consider foregoing the term altogether.

4. Trilemma Outcome

As suggested by "trilemma", we have three options. As an actualist, foregoing the rhetorical device is the least favorable route to choose. As a LMR, foregoing the indexical is the least favorable. If we decide to omit the term altogether, we commit ourselves to a possibility where either actual has any meaning. In other words, our world is merely possible and not actual in any sense. Of these three options, foregoing the term seems to be the most undesirable. However, one of the options allows for us to accept, indirectly, both LMR and Actualism. The details of this option are disclosed in §II. The details argue that the ontology of the LMR subsumes the ontology of the actualist.

Our first option of the trilemma is to accept actual as a rhetorical device. This option is the innate usage for the actualist. Even though C&H argue that this usage is redundant, there are worse outcomes; accepting this option and risking redundancy seems like a negligible price to pay. Paying this negligible price would commit us to Actualism and reject LMR. However, the main problem with this option, Bricker asserts, is that there are no *a priori* grounds to accept that everything is actual.[27] Although *a priori* arguments are lacking, the possibility of the actual world as the actualist sees it, remains open-ended. We have grounds for accepting the rhetorical device.

Our second option is to accept the actual operator as an indexical. The indexical is the

---

26   Bricker 2006. p56.
27   Bricker 2006. p56.

innate usage for the LMR. Without possible worlds, there would be no reason for this usage. Now even though the topic of parallel universes is used in fiction novels, movies, Sci-Fi, etc., accepting this usage would commit us to an actuality where we would be one of many novels, movies, and/or Sci-Fi possibilities - an actuality where we are one of a multitude. Being amidst a multitude would confront us with corollary arguments about the significance of life amongst an infinitude of lives. The price we pay for the indexical usage is ambivalence. Bricker's issue with this option is that there is more for an individual to know than if they are 'actual' at their world.[28] Even though Bricker rejects LMR, his concern for LMR has more to do with the overall theory rather than the utility of 'actual' as an indexical.[29] We have grounds for accepting the indexical. Our third option is to omit the operator altogether. If we accept this option, we may default towards LMR anyway. Omitting the operator entirely, for the actualist, would mean we would be unable to discern the actual world from a possible world. Without actuality, the criteria to differentiate our actual world from a multiplicity of possible worlds would be improbable, and we return to the issue of The Skepticism of Actuality. Even though this would solve the actualist's problem of redundancy they would have to trade their actuality in for possibility.

Omitting the operator entirely, for the LMR, changes little. The denizens of all the possible worlds would remain under the impression their world is the one world with a characteristic that separates them from the rest. The difference between the second and third option is the difference between concrete and abstract. Lewis claims possible worlds of his doctrine are all equally concrete; however, reluctantly so, because the distinction is unclear to him.[30] Regardless of Lewis' confusion, we may ignore the distinction between abstracta and concreta and forego the complexities of the third option if we only accept concreta.

II. LMR SUBSUMES ACTUALISM:

Traditionally, the debate between Actualism and Possibilism was intended to distinguish between actual and possible entities; each doctrine supporting their respective entities. The main distinction is between what are actual versus possible entities. The best way to observe this distinction is by studying their ontologies. Ontology is the study of what there is, or what exists. Inside the Actualist's ontology exists only actual entities. Within the Possibilist's ontology, possible entities exist – bearing in mind actual entities can also be possible entities. A superficial understanding of these theories may lead one to believe that the two doctrines are different doctrines because of what

---

28   Bricker 2006. p56, 69.

29   Bricker 2006. p69.

30   Lewis 1986. p81.

is considered in their ontologies. However, because of LMR, the distinction becomes less straightforward. Because even though the LMR is a doctrine under the umbrella of Possibilism, everything in the LMR ontology is 'actual' - actual in the indexical sense (refer to §I.2). §II in its entirety argues that the ontology of the LMR subsumes the actualist's ontology. To fully understand this claim, §II.1 begins with a brief exposition of both doctrines showing how they are different. Following the exposition, §II.2, discusses the role of States of Affairs and possible worlds and how the LMR ontology subsumes the actualist's. Lastly, §II.3 summarizes the conclusion of the section.

1. Distinction between Actualism & LMR

The main difference between the two doctrines were actual versus possible entities. Actualism is the idea that only actual entities exist. Meaning, zero nonactual possibilities exist. When we talk about dogs, cats, rabbits, etc., there are actual entities that instantiate those objects (I.e. actual dogs, actual cats, etc.). However, when discussing topics of merely possible entities; unicorns, centaurs, alternative options that you could have chosen but decided otherwise, etc., there fails to be any actual entities instantiating those objects. In other words, the difference between actual entities and possible entities are truthmakers. Truthmakers are entities that instantiate true sentences or propositions. I.e. something in the world exists that makes the sentence or proposition true. The main idea of truthmakers is that all truths have truthmakers.[31]

"A dog in the bog."

The only way this proposition holds any truth is if there is a dog, a bog, and/or a dog is inside a bog. By extension, for Actualism, sentences or propositions must be instantiated by actual entities if they are to express any kind of truth. Now, the actualist would agree that it could have been possible for merely possible entities (centaurs, unicorns, pots of gold at the ends of rainbows, etc.) to exist in the way that dogs, cats, and horses do, but for whatever reason those objects never actualized in the same way.[32] The way they actualize will be visted in §II.2, for now, what's important is that the actualist rejects nonactual possible entities in their ontology.

Contrastingly, LMR accepts nonactual possible entities. LMR, an idea of possible worlds, asserts, that every way a world can exist, exists such a world. There exists a world where you are writing this thesis and I'm reading it. There exists a world where this thesis is being written in a different language. There exists a world where you are writing this thesis and I'm reading it in a different language. The number of worlds that exist is vast enough that for every proposition there exists a set of possible worlds where that

---

31   Ney 2014. p152.

32   Menzel 2021.

proposition holds true at those worlds.[33] Meaning, truthmakers for the LMR-ist are 'actual' entities in a set of possible worlds that instantiate the entities of a given sentence/proposition. They can also exist in other sets of possible worlds. Example:

"Sasquatch ought to watch his watch."

For the LMR-ist, there exists a set of possible worlds where there is a sasquatch. Amongst the set of worlds where a sasquatch exists, there exists an instance in which it is wearing a watch. More specifically, amidst the set of worlds where sasquatch exists and wears a watch, it must be vigilant if it wishes to maintain possession of its watch. Therefore, the truthmakers of the proposition above, is an actual vigilant, watch-wearing sasquatch that exists in some possible world and has a timepiece that requires security, that gives this proposition a truth value. Intuitively, we may begin to see where the two theories begin to diverge - the actualist rejects nonactual possible entities while the LMR accepts them.

2. States of Affairs (SoA) vs. Possible Worlds

Even though the two ideologies approach nonactual possible entities differently, the differentiation acts as more of a boundary to distinguish between Actualism and LMR. Because one claim is that LMR ontologically subsumes Actualism, the distinctive approaches to nonactual possibilities are required to distinguish when the subsumption stops and begins. Without a proper understanding of this distinction, it would be reasonable to think that the two doctrines are unrelated. As reasonable as that may be, this section provides the proper understanding needed to understand this ontological subsumption through the idea of possible worlds.

A possible world for the actualist is a complete SoA. SoA are matters that obtain or fail to obtain.[34] For example, if you woke up this morning and had toast instead of cereal, the SoA where you had toast obtained, while the SoA where you had cereal failed to obtain. SoA that obtain are actual, while those that fail to obtain (some unobtainable) are possible. The difference between an ordinary SoA and a possible world is the property of maximality. Any maximal SoA is a SoA that precludes or includes any other SoA - that is, any SoA complements all other SoA.[35] A SoA that lacks maximality, or an ordinary SoA could be confused with a proposition. The difference between a SoA (complete or otherwise) and a proposition is that a proposition has a truth value – that is, propositions are either true or false, a property SoA lack. However, Plantinga agrees that a proposition can be true in an actual SoA. Therefore, a SoA, if it is actual, can have true propositions, and by extension, equates to a SoA that has the property of being

---

33   Lewis 1986. p53.
34   Plantinga 2008. p144.
35   Plantinga 2008. p144.

true. Similarly, the set of true propositions of the actual world, or the actual world with the set of propositions that make the actual world true, are the set of propositions that are only true in the actual world.[36] Meaning, the actual world is aligned with a set of true propositions of the actual world, giving the actual world a truth value. Therefore, the actual world as the actualist sees it, is true because the actual world is a set of propositions that are true. Now, for an actualist, the actual world is a possible SoA. Possible SoA are abstract objects. An abstract object is an object we consider, psychologically, without some of its features.[37] Abstract objects are also inert. Consider numbers, for example 3. Without any objects to instantiate, can you pick up 3? Can you smell 3? How about, taste it, hear it, and see it? If you have three tacos, you could certainly smell, taste, and touch [3] tacos, but the focus of the example is 3. Other than identifying tacos in example, 3 has no causal relationship to the tacos, it is causally inert. So psychologically, without physically possessing the materiality of any object, we can hold within the mind the number instantiated by the number of objects. Likewise, a possible SoA, which includes but is not limited to the actual world, can be held in the mind as an abstract object. In shorter words, the actualist suggests that the actual world is an abstract object.[38] Therefore, the abstract object that the actualist claims to be the actual world must be an abstraction of an actual object.

On the other hand, for the LMR-ist, a possible world is a proposition. More specifically, a proposition is a set of possible worlds. For every proposition is a set of possible worlds where the proposition holds true at the set of possible worlds.[39] Additionally, of the set of possible worlds that instantiate a proposition; each and every one is spatiotemporally isolated, complete, and concrete.[40] Lastly, another relevant key feature is that although these possible worlds are like an actualist's complete SoA, every world is unrelated to any other world; nothing that happens at one world can affect another.[41] Even though the LMR-ist believes all worlds are concrete, possible worlds exist in a way much like the actualist's actual world - unable to causally effect another world; inert. So, if the actualist claims the actual world is an abstract object and the LMR-ist claims all worlds are concrete, but both actual world of the actualist and concrete worlds of LMR are causally inert in relation to all other worlds (concrete or otherwise), how might one distinguish between a possible concrete world and an actual abstract world?

Ontologically speaking, this world; our world, has all the constituents of a world

---

36   Plantinga 2008. p144-145.
37   Ney 2014. p64.
38   Plantinga 2008. p144.
39   Lewis 1986. p53.
40   Lewis 1986. p2.
41   Lewis 1986. p2.

that could be the actual world of an actualist or an 'actual', possible world of a LMR-ist. Everything in this world exists; that is, everything in this world is actual and/or 'actual'. This keyboard I type these sentences on, the screen you read these sentences I've typed, the monitors by which we exercise our respective roles, etc., all these objects are actual and exist. The question then becomes, "Exist in what sense?" Do these objects exist in the possible-concrete-world-sense that the LMR-ist claims possible worlds to exist? Or do they exist in the way of the abstract-object actual-world sense that the actualist claims the actual world exists? In either sense, what we can confirm is that whatever is in the ontology of the actualist can be found in the ontology of the LMR-ist. Simply by nature of actual and possible entities; all actual entities are possible entities, and in an ontology of possibility (LMR) all actual possibilities can be found (Actualism). More succinctly, the ontology of the actualist's is subsumed by the LMR's.

3. Accepting LMR subsumes Actualism:

In summary, §II.1 articulates that the chief difference between the two doctrines is the prevalence of nonactual possible entities; the actualist rejects while the LMR accepts them. §II.2 details how both doctrines accept possible worlds in their ontologies. We know that the ontology of the LMR is abundant, if not infinite; at a minimum, sufficient to account for the ontology of the actualist's. If we know that the ontology of the LMR is sufficient to account for the ontology of the actualist, then there's enough reason to suspect that the actualist's actual world is an abstraction of an 'actual' LMR possible world. Our world is a possible world, which could mean we are a possible world of LMR, implying that our world is also concrete. Our world is also a possible SoA, which means we could be a SoA that resembles the actualist's actual world, implying our SoA is also abstract. In both instances our world is inert relative to other worlds. A LMR-ist possible world, by doctrine, is inert (spatiotemporally isolated). An actualist's SoA, by definition of abstract, is inert. However, like the previous taco example, abstract objects are better observed when instantiated by concrete objects. We could argue our world is both the abstract SoA that is the actualist's actual world and an 'actual' LMR possible world. We can make this claim because the aperture in which you view the world can be slightly different from the aperture in which I see it; however, we exist in the same world. Even though we subsist in the same world, because we see the world slightly differently, the abstractions we hold in our mind of our world may vary. In theory, we're idealizing the same abstraction. In actuality, just as you may have never been to my elementary school in Liberal, Kansas and I have never been to an elementary school in Boston, Massachusetts, our respective abstractions will slightly differ. Fortunately for us, LMR has a set of possible worlds that instantiates every possible world that can exist. Meaning, the slightly different, but same abstractions we imagine in our minds, is instantiated by a concrete possible world. Furthermore, there is not just one world,

there is a set of worlds where both elementary schools exist in the way that we remember them and they exist the way for us as they do for the other, even if you have never seen mine, and I yours.

III. ADVANTAGE LMR

Outside the philosophy room, there is less structure, clarity, consistency, etc., during discussions of modal concepts, specifically when possible worlds come into play. Therefore, "It is one of the tasks of a philosophical theory of modality to give a systematic and unified account of this multiplicity of modal concepts.".[42] So rather than haphazardly regurgitating interesting quotes and ideas from random authors and philosophers without any governing principles, Actualism and LMR are formulaic and consistent doctrines that satisfy modal concepts. Inside the philosophy room, theories like Actualism and LMR provide consistency when talking about possible worlds. While Actualism satisfies modal claims with actual entities, LMR answers them with possible ones – which encompasses actual entities. §III in its entirety addresses the methodologies of both doctrines and how they satisfy modal claims. §III.1 details the methods of each doctrines' strategy for modal concepts; counterfactuals and essences. §III.2 discusses how counterfactuals have an advantage over essences. Lastly, §III.3 concludes with a preference for LMR over Actualism and why it is more advantageous.

1. Counterfactuals and Essences

Modal claims express some sort of fact. Depending on the type of modal claim it will have some modifier prefacing the claim; i.e., 'it could have been that…', 'it is necessary that…', 'it is possible that…", etc. The purpose of these claims is to verify the logic that determines the validity of an argument.[43] Valid arguments are when the premises support the conclusion, whereas invalid arguments lack this support. A LMR-ist uses counterfactuals to satisfy modal claims, while the actualist uses essences. For Lewis, counterfactuals are an invitation to consider possibilities at a given possible world.[44] Consider a series of events, A, happens at world, WA, but series of events B, C, and D, were also likely outcomes, Lewis considers worlds WB, WC, and WD as possible worlds where those alternative outcomes would have occurred; counterfactuals. Additionally, Lewis equates counterfactuals to causation.[45] Which means, the outcome of WA is caused by a series of events specific to WA, the alternative possible causes that could have happened are instantiated by a specific series of events relevant to their respective possible worlds (WB, WC, etc.). Imagine you wake up and have coffee for breakfast. After

---

42   Kment 2021.
43   Garson 2021.
44   Lewis 1986. p20-21.
45   Lewis 1986. p23.

you have your coffee, you may need to take a trip to the loo. Depending on where you're at, you may use the restroom at home or a public one. Two potential outcomes caused by the same prior event. Both outcomes happen. One at WA, the other; counterfactually, at WB. Although you exist in WA, drinking coffee caused a fork; WB, where you went to the other restroom instead. Which also implies, another possible way a world could exist is, WC, the world which coffee was never an option, nor a trip to the loo. Briefly, every cause produces potential outcomes and possible worlds where alternative outcomes occur, counterfactually.

Now, even though an actualist may use counterfactuals; they're instances that could have occurred but failed to obtain. Therefore, counterfactuals, for the actualist, are considered *possibilia* rather than actual instances. WA, for the actualist is the only world that exists - WB and WC are *possibilia*. For the LMR-ist, WA is an 'actual' possibility that exists – WB and WC are spatiotemporally isolated from WA; however, also both exist in the same way.

Contrasting the counterfactual method, the actualist utilizes actual entities to satisfy modal claims. However, there are modal claims that for an actualist, fail to obtain - claims that incorporate nonactual possibilities. As an actualist, essences are used to satisfy modal claims containing nonactual possible entities. An essence is a "certain quality, singular and incommunicable to any other subsistent".[46] A unicorn has a specific essence that is singular and incommunicable to any other subsistent - the only thing that has the essence of a unicorn is the unicorn. Although unicorns fail to obtain in the actual world, there is a SoA in which they do obtain. In the SoA that unicorns do obtain, so does their essence.[47] So even though unicorns fail to obtain in the actual world, the ontology of the actualist still includes the essence of them because the essence exists in some SoA. Thus, allowing the actualist to address modal claims containing nonactual possible entities. In example, both doctrines have different ways of satisfying this modal claim:

"It's possible that in Sasquatch-town, yetis commute on unicorns."

For the actualist, there are essences of all the constituents above, yetis, unicorns, etc.. Now, even though those constituents failed to obtain in the actual world, they do exist in some possible SoA. So, to clarify, the essences of those constituents exist, but the material that would have instantiated those constituents in the actualist's world remain possibilia. Contrastingly, a LMR-ist asserts that there is a set of possible worlds that satisfies that claim. At the worlds where there exists a town named after Sasquatch(es), and in those towns there exists yetis that commute on unicorns, that proposition holds true. Additionally, it 'actually' exists in the same way the other concrete possible worlds and the actual world of the actualist exists; causally inert from any other possible SoA/

---

46   Plantinga 2008. p149.
47   Plantinga 2008. p155-160.

possible world.

2. Counterfactuals over Essences

Both doctrines meet the challenge of being a philosophical theory that provides a formulaic and consistent account of addressing modal concepts. However, the methodology used by the LMR-ist has two advantages over the actualist's methodology: 1) The LMR-ist approach is simple and straightforward. 2) Counterfactuals can encompass essences – we can include both methodologies. To observe the first advantage, suppose we imagine there was a cause that redirected monkeys from evolving into humans, and instead, into a colony of bigfoots; bigfeet. This evolutionary supposition for the LMR-ist is a counterfactual. The cause and likelihood are less important than the idea that the counterfactual is equally 'actual' as our possible world. Meaning, along with our actual possible world, there exists a world spatiotemporally isolated, and acausal where bigfeet thrive. For the LMR-ist the theory is straightforward, there exists a world where a colony of bigfeet exist because they evolved from monkeys, or however else they may have come into existence. The truthmakers are any world where actual bigfeet subsist.

Alternatively, essences for the actualist, are a roundabout method of addressing modal claims. For instance, in some possible SoA, included, but not limited to the actual world, exists the essence of a bigfoot. Because the essence of bigfoot exists in some SoA, the actualist acknowledges bigfoot in their ontology, and therefore can address modal claims concerning bigfoot as more than just possibilia. Even though in the actual world bigfoot may have failed to obtain, the essence of bigfoot exists in some SoA. So, from the actualist perspective, the truthmaker for bigfoot is its essence, and that essence satisfies any modal claim pertaining to bigfoot. In other words, for the actualist, bigfoot is more than just mere possibilia. Bigfoot is an actual entity whose essence manifests in some SoA. Although Bigfoot's essence may have obtain in another SoA, it fails to obtain, or is unobtainable in the possible SoA that is the actualist's actual SoA. Even though we can conceptualize the actualist's methods on how bigfoot exists, the LMR-ist's straightforward approach has the advantage of simplicity.

Additional to the straightforward approach, counterfactuals as used by the LMR-ist, can also include essences. Although they would possess a different name, the essential idea would remain the same. Recall the second advantage 2) counterfactuals could encompass essences. Essences for the LMR-ist are understood as the character of a counterfactual. Each counterfactual, like an essence, has an identifiable characteristic specific to that series of events; however, in the context of counterfactuals, it's the specific sequence of events that uniquely identifies each concatenation. Just as a single electron identifies a hydrogen atom, two electrons; helium, three electrons; lithium, etc., every counterfactual has a particular series of events that is specific to its unique sequence; providing each specific counterfactual with an identifiable characteristic distinguishing

it from every other. In this light, we can consider counterfactuals of having their own essence – at the very least, "...a quality incommunicable to any other...".[48] So not only is the LMR-ist methodology direct and clear, but it also offers how the actualist's methodology can be incorporated - no need for debates or deciding between two methods.

3. Accepting LMR over Actualism

As previously stated, modal claims express some sort of fact. Whether counterfactually, or through essences, our fact of existence is a modal claim. From the actualist perspective, there exists a SoA with combination of essences that includes everything and anything pertaining to the SoA that exemplifies the fact of the actual world as we understand it to be. From the LMR-ist perspective, there are sets of possible worlds that not only instantiate, the actualist's actual world, but also, instantiate a virtually infinite number of similar combinations accounting for every possibility. The historical polarity between these types of doctrines suggests we can only have one or the other. However, the focus of this thesis, in tandem with contemporary research, suggests a non-mutually exclusive outcome that considers the simultaneity of both doctrines. Of course, no doctrine is perfect. This thesis would be tendentious at best if the pitfalls of LMR were left undisclosed. Surely with an infinite number of possibilities occurring something is bound to go wrong. Short answer, yes. However, the problem is a common one within the discipline of Philosophy rather than unique to LMR. If there are an infinite number of realities just like ours, and others where there are less problems, and those numbers reflect others with more problems, what sort of meaning would our 'actual' existence have? With so much to consider, you could see how one could spiral into a metaphysical nihilism – aimlessly wondering about, because we could potentially just be another rendition of a forgotten idea in an unknown location where any significance worth having is inapplicable to our inert existence; concealed in a perpetual purgatory to reenact an eternity, endlessly. A couple ways to answer this question: 1) those problems are inaccessible to us and 2) this is the advantage to concreteness. Firstly, as a LMR-ist, this is an awkward question. You could ask "How much does four weigh?", "What's it like to taste six?", "How many can an eight feed?", etc., and the words would make sense, but the answers are inapplicable. For a LMR-ist, because each world is spatiotemporally isolated, those answers are inaccessible to us. For instance, the sets of worlds where Hilary Clinton was elected president, only those worlds can address those phenomena regarding President Clinton. In the set of worlds we exist in, whether Hilary was reelected in the 2020 election, makes sense to answer; however inapplicable to our situation. Any and all things regarding Hilary's reelectedness is inaccessible to us. Secondly, "How much does four weigh?", "What does six taste like?", etc., highlights the advantage of concreteness.

---

48 Plantinga 2008. p149.

Rather than just asking about numbers, apply "tacos" to those questions. "How much do four tacos weigh?", "What does six tacos taste like?" etc., there are accessible answers to those questions when there are concrete objects instantiating. Imagine a memory and/or a dream. Really picture the idea in your mind's eye. At some point you're unable to pinpoint the beginning. Similarly, you probably are unable to locate the end, or it abruptly ends with an open-ended conclusion. Brain images spontaneously manifest without beginning or end - in all cases, ideas remain abstract. Compared to the concrete reality LMR argues for, you have the beginning of your life that begins at your birth and ends at your death. Otherwise, one can argue we're just living out a dream, memory, idea etc.. So even though LMR raises some questions, there are answers.

As argued, accepting LMR over Actualism posits two advantages. The first advantage was directness and clarity - for any possible SoA that is the actualist's view of our actual world, that SoA exists concretely. No need to explain essences, possible SoA, how sometimes some SoA obtain, and/or how unobtainable as some might be, they still exist in some fashion. Directly and clearly, all possibilities 'actually' exist. Secondly, even if we refuse to reject Actualism, or we prefer essences, or any other reason why we might consider Actualism over LMR, essences are characteristics of counterfactuals. Choosing between either doctrine is unnecessary. We can accept the utility of both doctrines without accepting any discontinuities. Lastly, like all doctrines, LMR has its issues, but it also addresses those issues without all the circles.

CONCLUSION

The central focus of this thesis was that Actualism and LMR are not mutually exclusive. Even though Actualism is a theory consisting of only actual entities and LMR a theory consisting of possible entities, as argued, it's possible the two doctrines are concomitant. With the unorthodox way of utilizing the term actual, this non-mutually exclusive relationship is represented by the LMR-ist's ontology subsuming the actualist's.

An argument could be made that even though the ontological relationship suggests a concomitant relationship; it's merely suggestive, and not necessarily concomitant. The abstract object that is the actual world of the actualist, *could be* instantiated by a LMR concrete possible world, it doesn't necessarily mean that it is. Ultimately, both worlds (actual world of the actualist and 'actual' world of the LMR-ist) are causally inert, so whether they are or aren't the same world, because they're both either causally inert or spatiotemporally isolated, the absolute actual world (arguably, the world that we exist in) resembles the quantum superposition of Schrödinger's cat; both concrete and abstract until an observer lifts the proverbial box and reveals the truth of our nature - an observation privy only to an entity beyond the system. Until then, our actual world is inconclusively abstract and concrete; suggesting, minimally, that Actualism and LMR

are not necessarily mutually exclusive.

Another objection regarding one of the claims of this thesis, is the necessity of the term actual. An argument can be made that omitting the term actual doesn't negate actuality. I.e. some cultures lack the words for specific shades of colors – fuchsia. Fuchsia is a shade of pink that has a tinge of red and purple. Some cultures lack the word for pink, let alone a shade of pink. That doesn't mean that fuchsia doesn't exist. This argument can be superimposed onto the term actual. Just because we omit the term actual does not mean actuality doesn't exist. "Fair enough.", is my retort. However, like the color fuchsia, "How is one supposed to discern pink from fuchsia or any other shade of pink for that matter?" More relevantly, "How does one discern an actual world from a possible world?". If we presuppose Actualism, and our one world is the only world and it is actual, what specific criteria does it possess that discerns it from all the other possible worlds that could be the actual world? "Simple.", an actualist might retort, "It's the only one that exists. There exist no other actual worlds." And that would be no problem if LMR wasn't also a theory. Because even though the actual world from the actualist perspective believes that there is only one actual world, the nature of that world is causally inert. That 'only one actual world' has the same description of a LMR possible world, and those worlds are not only 'actual' they're concrete too. So possibly, our world holds the quantum superposition of being both the actual world of the actualist and an 'actual' world of a LMR-ist. Otherwise, we must visit the abstract vs concrete debate. Regardless of the outcome of that debate, the fact of the matter is, without the term actual, although there may still be an actuality, the nature of actuality would be abstruse to say the least.

In conclusion, Bricker argues, we're unaware of the nature of our reality. Lewis; the quintessential possibilist suggests, there is nothing to fear - all possibilities 'actually' exist, concretely too. Plantinga; an actualist suggests, we're the one and only actual, abstract SoA. Both definitions, be it possible world or SoA, are causally inert to any other world. Nothing of said object(s) can go beyond and nothing outside can infiltrate. Because of the term actual, we can accept LMR and encompass the ontology of Actualism and therefore skip any complicated debates whilst reaping the advantages of both doctrines. Some may argue that the conclusion is merely a suggestion and needs more compulsion. Be that as it may, at a minimum, Actualism and LMR are undeniably non-mutually exclusive – Schrödinger's cat is also TBD, but both dead and alive, nonetheless. Others may argue that omission of the term actual changes nothing; we exist regardless of how the term is or is not used. Perhaps that cohort may be onto something. That cohort must then confront the nature of our reality, and as we know from the beginning of this thesis, Bricker leads that cohort, and look where that got us.

❦

BIBLIOGRAPHY

Bricker, Phillop. "Absolute Actuality and the Plurality of Worlds." Philosophical Perspectives, vol. 20, no 1, Dec. 2006, pp. 41-76. DOI.org(Crossref), https://doi.org/10.1111/j.1520- 8583.2006.00102.x.

Crossley, John & Humberstone, Lloyd. (1977). The Logic of "actually". Reports on Mathematical Logic. 8.

Davies, Martin, and Lloyd Humberstone. "Two Notions of Necessity." Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, vol. 38, no. 1, Springer, 1980, pp. 1–30, http://www.jstor.org/stable/4319391.

Donnellan, Keith S. "Speaking of Nothing." The Philosophical Review, vol. 83. no. 1, [Duke University Press, Philosophical Review], 1974, pp. 3-31,

https://doi.org/10.2307/2183871.

Garson, James, "Modal Logic", The Stanford Encyclopedia of Philosophy (Summer 2021 Edition), Edward N. Zalta (ed.),

<https://plato.stanford.edu/archives/sum2021/entries/logic-modal>.

Hanson, William H. "Actuality, Necessity, and Logical Truth." Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, vol. 130, no. 3, Springer, 2006, pp. 437–59, http://www.jstor.org/stable/4321811.

Humberstone, Lloyd. The Connectives, MIT Press, 2011. ProQuest Ebook Central, https://ebookcentral-proquest

com.ezproxy.otago.ac.nz/lib/otago/detail.action?docID=3339269.

Kaplan, David. "On the Logic of Demonstratives." Journal of Philosophical Logic, vol. 8, no. 1, Springer, 1979, pp. 81–98, http://www.jstor.org/stable/30227158.

Kment, Boris, "Varieties of Modality", The Stanford Encyclopedia of Philosophy (Spring 2021 Edition), Edward N. Zalta (ed.),

<https://plato.stanford.edu/archives/spr2021/entries/modality-varieties/>.

Kripke, Saul A. Naming and Necessity. Rev. and enl. Ed., Paperback ed., Repr, Blackwell, 1990.

Lewis, David K. On the Plurality of Worlds. B. Blackwell, 1986.

Ney, Alyssa. Metaphysics: An Introduction. Routledge. 1st edition. 2014.

Menzel, Christopher, "Actualism", The Stanford Encyclopedia of Philosophy (Fall 2021 Edition), Edward N. Zalta (ed.),

<https://plato.stanford.edu/archives/fall2021/entries/actualism/>.

Menzies, Peter and Helen Beebee, "Counterfactual Theories of Causation", The Stanford Encyclopedia of Philosophy (Winter 2020 Edition), Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/win2020/entries/causation-counterfactual/>.

Plantinga, Alvin. 'Actualism and Possible Worlds'. Theoria, vol. 42, no. 1–3, Feb. 2008, pp. 139–60. DOI.org (Crossref), https://doi.org/10.1111/j.1755-2567.1976.tb00681.x.

Plantinga, Alvin. The Nature of Necessity. Clarendon Press, 1974.

Robert Stalnaker, I—Robert Stalnaker, Aristotelian Society Supplementary Volume, Volume 75, Issue 1, 1 July 2001, Pages 141–156, https://doi.org/10.1111/1467-8349.00083.

Rodriguez-Pereyra, Gonzalo. Leibniz: Discourse on Metaphysics. Oxford, 2020;

onlineedn, Oxford Academic, 20 Feb.2020),

https://doi.org/10.1093/oso/9780198829041.001.0001, accessed 18 Sept. 2022.

Seth Yalcin, Actually, Actually, Analysis, Volume 75, Issue 2, April 2015, Pages 185–191, https://doi.org/10.1093/analys/anv021

Starr, William, "Counterfactuals", The Stanford Encyclopedia of Philosophy (Summer 2021 Edition), Edward N. Zalta (ed.),

<https://plato.stanford.edu/archives/sum2021/entries/counterfactuals/>.

Williamson, T. 'Necessitism, Contingentism, and Plural Quantification'. Mind, vol. 119, no. 475, July 2010, pp. 657–748. DOI.org (Crossref), https://doi.org/10.1093/mind/fzq042.

Zalta, Edward N. "Logical and Analytic Truths That Are Not Necessary." The Journal of Philosophy, vol. 85, no. 2, Journal of Philosophy, Inc., 1988, pp. 57–74,

https://doi.org/10.2307/2026992.

AGAINST THE GUISE OF THE GOOD

by BRIGHT DUA-ANSAH

University of Tartu

ABSTRACT ‖ *The "Guise of the Good" thesis explains the nature of intentional action as one that is aimed at accomplishing some good. This makes sense as an explanation of why people do what they do, since we are less likely to deliberately act upon intentions in which we see no good. Francesco Orsi has argued that an agent's perception of an action as good is a necessary reason for its being carried out. I show that Orsi's thesis is merely a contingent explanation of intentional human action, and hence that the Guise of the Good thesis is false.*

People have the capacity for intentional action. When they do so, we like to think that they are aware of what they are doing, what they aim to achieve, and sometimes what they are trying to avoid. For instance, the purpose of this essay is to explore a claim. For this action to be intentional, I must believe chiefly that the purpose for which I am doing it is worthwhile (else, why do it?). This rudimentary assessment of my action is simply what the Guise of the Good thesis (hereafter, *GG*) is all about. In concise terms, when we act intentionally, we do it because we think it is good.[1] Expressed differently, if

---

1   In this discourse, the ethical debate of what constitutes *goodness*, rightness, and the summum bonum, inter alia, are all taken as settled. My focus is not to discuss the ontology of goodness.

we do not find something to be good, we will not do it—and even if we did it, our action would not count as intentional action (Orsi, 2015). Intentional actions are done under a guise that is believed to be good, so to speak.

The question which I propose to answer in this essay is briefly this: when we act intentionally, are we always acting under the Guise of the Good? In what follows, I will suggest that the answer is no.

In Section 1, I give a more precise account of *GG*, including prior versions that have contributed to its present formulation. I end the section by noting some problems with the thesis such as its limited scope, the apparent existence of some intentional actions that are not done under the Guise of the Good, and the kinds of intentions which can make an agent act in the contrary to what is good. Given these problems, we have reason to believe that the *GG* thesis may be false.

In Section 2, I discuss which part of *GG* is most likely to be mistaken. I content that *GG* is not necessary for explaining intentional action. Following this approach, I show why this false conclusion is a result of false premises: in particular, *GG* gives too much authority to the agent. The general approach of the section inverts the argumentative strategy of the *GG* thesis. I highlight some attempts that have been or can be made at reorienting the thesis against my objections. For instance, one response attempts expanding the scope of the Guise of the Good to include actions that are not considered by agents as good.

In Section 3, I explore the relationship between intentions and actions through three thought experiments. I believe that these examples illustrate deep problems with *GG* that can be neither ignored nor explained away. Through my discussion of these examples, I hope to contribute substantial evidence against the *GG* thesis.

SECTION I

1. What is the Guise of the Good?

In order to clarify the *GG*'s commitments, I will first elaborate on some key concepts that the thesis employs both explicitly and implicitly. In action theory, a rational agent is one who has the capacity to understand the "why" of her actions. In requiring understanding for intentional action, *GG* thus restricts its scope to rational agents. What makes an act rational is its being considered admissible by the agent in question—though not necessarily by everyone. For instance, Leonidas' defending Sparta to the death in battle might not make sense to you or me, but if he believed he was doing something intelligible, his actions were rational. [2]

---

2   I refer to the Battle of Thermopylae (Θερμοπύλαι), 480 BCE, where the Spartan king
      Leonidas (along with some Lacedemonians and a small contingent of Thespians) attempted
      to ward off the Persian military campaign under Xerxes I (son of Darius I) at the hot gates,

Further, and more critical to *GG*, is the notion of *reasons*. In her seminal work Intentions (1963), G. E. M. Anscombe suggests that asking for the *intentions* of an agent A is demanding to know from A the "why" of her action. For instance, if an agent turns left on a highway to a nearby motel while driving, a mechanic can offer a partial explanation of why she turned and drove in that direction. They might discuss the effect of turning the steering wheel counter-clockwise, the corresponding effect on the rack and pinion system of the automobile, and so on. Likewise, a neuroscientist can talk about the various neural responses involved in making such a choice. But all this will not be adequate when we ask the agent her intention for turning leftward. She might simply tell you she saw a motel and that wanting to catch some rest, relax, and get some tasty food, she turned there. None of the technical explanations of the mechanic or neuroscientist are helpful or even relevant when we asked for her intentions. Rather, the agent owes us a *justificatory reason*, or more importantly an *explanatory reason*, for turning left.

An important point of note in the historical antecedents of action theory is the notion of desires and wants. In his cosmic ordering of plant and animal life, Aristotle suggests that the locomotion of animals is due to their basic wants and need to survive. He writes,

> For, where there is sensation, there is also pleasure and pain: and, where there are, desire also must of necessity be present.[3]

On this account, no elaborate system of evaluation is required to act. All that is required is that one possesses the rudimentary senses that respond to stimuli positively and negatively.[4] Like Aristotle, David Hume presents a thorough exposition on the distinction between desires and beliefs. Desires, he claims, are non-cognitive and can motivate action, while beliefs are cognitive but cannot motivate action. According to this view, *Motivational Humeanism*, beliefs are cognitive in that they, unlike desires, involve propositions about the world that can be false or true.[5] Note, however, that our theoretical evaluations of *the good* and our desires may nevertheless coincide, as Kant argues in the *Critique* and *Groundworks*.[6]

---

which refers to a narrow mountain pass strategic to managing the overwhelming Persian army. Defeat seemed inevitable, but Leonidas stood his ground.

3   Aristotle, *De Anima* III. 1925: 8,9 (413b 22 – 414a 1).

4   Precisely, Hume (1878) THN 2.3.3

5   Philosophers like to draw a distinction between cognitivism and non-cognitivism. The implication is that, depending on which school one belongs to, one can argue for the validity (or lack thereof) of objective moral notions. If one is a cognitivist, then one can argue that moral facts can be discovered by the conscious individual. If one is a non-cognitivist, then there are no objective moral notions; instead, one is ultimately talking about desires and tastes in valuing. See (Ayer, 1936); (Hare, 1952); (Mackie, 1977); (Blackburn, 1984); (Gibbard., 1990); (Shafer-Landau, 1998).

6   See (Moore, 1903); (Hare, 1952); (Mackie, 1977); (McDowell, 1985); (Korsgaard, Skepticism About Practical Reason, 1986); (Smith, 1994); (Dancy, 2000); (Orsi, The Guise

The *GG* theory promises to answer the question of how we act and on what motivations we do so. *GG* argues that any desire to act rationally must be formulated with a normative justification in mind. Consequently, *GG* is resolved that all intentional actions are done with good foreseen effects.[7]

The strength of the connection between the motivating and justificatory reasons for acting varies according to different authors. For Scheffler (2011), it may be an aesthetic, moral or pragmatic consideration. For Scanlon (1998, 2011) and Korsgaard (2009), there exists an evaluative content (or *rationalism*, as Kantians prefer to call it) entailed in acting intentionally. Like Orsi, I formulate my views according to the latter interpretation.[8]

The *GG* thesis can be argued for in the manner below:

P1   If action *x* is intentional, then agent *A* understands why she did *x*.

P2   If *A* understands why she did *x*, then *A* has justificatory reasons for doing *x*.

P3   If *A* has justificatory reasons for doing *x*, then *A* sees *x* to be done under *GG*.

C   Therefore, if an action *x* is intentional, then *A* sees *x* to be done under *GG*.[9]

In what follows, we consider some problems with *GG* as presented in the formal argument above.

2. Problems Confronting the GG

GG does not merely generalize that all human intentions are targeted at *the good*. As my interlocutor, Orsi, proffers, "In any case, [*GG*] is held to be more than a contingent generalization about human beings like 'most humans act under the guise of *the good*.'"[10] The relation, rather, is one of necessity: all human intentions must target *the good*. What this implies is that, granted one is acting intentionally, it follows without failure that they are acting under *GG*.

---

of the Good, forthcoming).

7   As Stocker (1979) posits this is a consequence of the "metaphysics of psychology" owing to Aquinas, Spinoza, Perry and even Sartre. Aquinas states it thus: *quidquid apetitur, apetitur sub specie boni*; whatever is desired, is desired under the guise of *the good* [*Summa Theologica* I-II: 1.6; 8.1, cited in Orsi (2015:714)].

8   This preference for the kantian interpretation in this article is that it is the more dominant interpretation among the *GG* theorists. It can be tracked back through Leibniz to Plato, who believed that people will do *the good* if they had knowledge of it because it is better for them anyway.

9   One finds versions of the *GG* in (Plato), Malebranche, etc. James Doyle has approached the *GG* from the epistemological standpoint. My simplified rendition reflects the core of the *GG* thesis in the literature I want to tackle. This final formulation comes after correspondence with Prof. Orsi who rejected my earlier formulations: I had understood his argument as establishing a biconditional relation between intentions and the *GG*.

10   Orsi, 2015: 715

Are the so-called intentional actions done by human beings, per se, or by rational agents? The difference between human beings and rational agents may be overlooked, but I contend that the distinction is one of substance. In this debate, it is human beings that we speak of; it is these same human beings that will act rationally or intelligibly. But *GG* does not specify whether it is humans beings spoken of or "rational agents." For instance, when we mention "rational agent", we can predicate some things of rational agents that we cannot do for "human beings," and the opposite is also true. Consider these expressions:

(a) *Respect every human being.*
(b) *Rational agents act rationally.*
(c) *Do not slap any human being.*

We do not think of the term "rational agent" and the term "human being" as interchangeable. Expression (b) is tautologous, whereas the expression "Human beings act rationally" may be objected to. This distinction in terms is not made by *GG*, which renders problematic its attempt to explain human action through recourse to the rational agent.

Furthermore, is it truly the case that *GG* is necessary for explaining intentional action as intrinsically tied to pursuing *the good*? Other theorists, like Setiya (2010), try to stretch the scope of *GG* such that intentional actions from other guises spill over into the domain of *GG*. For example, if I act from a desire for fame, we can explain my action by saying that I take fame to be good. Thus my act is motivated by *GG*. One could even argue, therefore, that the possibility of action under the guise of the bad does not contradict *GG* (Raz, 2016). Unfortunately, the ontology of intentions does not easily allow for this promiscuity. In addition, it intuitively seems that intentions can allow for actions that the agents themselves do not see as good. We will delve into this much more in Section 2.

Finally, my opposition to the *GG* can be tracked back into the second premise, P2, which states: *If A understands why she did x, then A has justificatory reasons for doing x*. One needs only to understand her reason for acting; whether A believes she is justified in doing so is a separate matter. It is also worth noting that no non-intentional action can serve as a counter-example to P1 regardless of whether the consequent of P1 (that *the agent A understands why she did x*) is true. It is my belief that P1 is set up this way to navigate the problems such as I have accentuated in the last three paragraphs. In what follows, I will address these more closely, beginning first with my opposition to *GG*, and then the other concerns I have raised.

SECTION 2

1. That the GG is false

I hold that as an account of intentional action, *GG* is false. Let us first consider a couple of broad objections to the assumptions on which the argument's relevancy depends and then move to discuss narrower concerns.

First, it is possible that there exists no intentional action, x, in the set of all actions that an agent, A, does in her universe. The very nature of *P1* is such that it will hold unless the antecedent (*if an action x is intentional*) is true and the consequent (*the agent A understands why she did x*) is false. All subsequent inferences from *P1* may still follow without impediment even if there is no intentional action.

Second, I would call into question whether *P2* coheres with experience. Let us consider the example given by Setiya (2010): a man runs from a burning house and you ask him why he ran. He can explain what he is doing and understand it all the same. But when you ask him if he believed he was justified in running, he may still answer no to that. His reasons may include that by running, he left his family in the house. But that does not take away the fact that his actions are intelligible, that he understands why he acted in a particular way, and that he ran for fear for his life. If *A* does not have justificatory reasons for doing *x*, it is not right to say that therefore *A* does not understand x. If A understands why she did *x*, it does not follow that *A* believes also that she is justified to do x. To have a reason for doing an intentional action *x* is to have an explanatory reason for why you did x; it does not necessarily entail a justificatory reason for doing *x*. So it seems that P2 may be false. I suggest in the following section that the reason for this lies in the ontology of intentions.

2. The Ontology of Intentions as Reasons

In Section 1, we learned the importance of intentions in action theory. *GG* argues that it is the nature of intentions to track and aim at what is good and worthwhile. The *GG* thesis posits a necessary relation between the Guise of the Good and all intentional action. This makes any intentional action under any other guise seem not only improbable (if at all it can be shown), but also strange to understand (Setiya, 2016). Yet, the ontology of intentions not only allows for this possibility but also generates a proliferation of actions done under other such guises.

The ontology of intentions is neutral with respect to the guises under which we act: it is neither for nor against any guise as such. Intentions are normative in nature and form the evaluative content of what motivates action. When we ask A: *Why did you save the baby?*, the response that she should give takes the form: *I did so because … .*, where the gap is filled with some normative account that should be her reasons. Of course, whether those reasons carry the argumentative weight needed to justify her actions is

another question, but that is not the point here; in order for her action to be intentional, she only needs to acknowledge she has some reasons for doing *x*. The problem here is that, when *A* decides to ϕ, that will constitute an intentional action *x* so long as *A* sees ϕ-ing to be worth doing (Raz, 2011).[11] Charles Ponzi would be acting under *GG*, and so would Martin Luther King Jr. In both cases intentional actions are done, according to *GG*, because each agent considers his actions worthwhile to do.

The aim of the last few paragraphs was to illuminate my suspicion that the ontology of intentions is not specific to any guise. This will explain why Ponzi and Dr. King can both be acting intentionally. The possibility of agents such as King and Ponzi raises suspicions about *GG*. How so?

The guise under which all intentional action is done (according to *the GG*) is motivated by *the good*. The good has a magnetic feature (Stevenson, 1937). As such, in terms of action theory, it attracts intentional action. We expect that only a limited class of actions are necessarily attracted by *the GG* in this way. Yet strangely enough, on the other hand, intentions have a strange feature of being agent-relative. What is good is only good insofar as the agent views it as good. In this regard, agency works as a *normative transformer*.[12] Neither the consequence of an action nor the Weltanschauung of the society in which the agent acts can determine action this way; even if they happen to influence action, neither of these can be considered the primary or necessary cause of action. Rather the values of the agent in question are the final arbiters of the worthwhileness of the action. Depending on their mood or psychological state, the same agent can briskly evaluate their action as worthwhile and done under the *GG* at time *t2* in opposition to a contrary evaluation at an earlier time, *t1*. At *t1*, I might find that a good cause justifies a war to me; at *t2*, I might find conversely that a good war justifies a cause to me, or that neither is justified at all.

In this *subsection 2.2*, we have identified that the ontology of intentions can allow for contradictory actions under *the GG* thesis. These normative grounds that motivate action must be stable enough to have the force of justification. *The GG* has a few possible responses to attempt to resolve these problems, such as:

---

11 See Ch. 2 Reasons: Explanatory and Normative. Raz also identifies this as a weakness of the thesis.

12 This term is borrowed from *Sex, Culture and Justice: The Limits of Choice* (Chambers, 2007). It is used to suggest that gender differences in workers' compensation are due to choices females make that coincidentally lead to their tendency to earn lesser wages than males. Hence, gender differences in compensation are not unjust since the difference is a result of choice. Applied to this discussion an action that may not be intentional because it is not justified by an agent, *A*, nonetheless may be intentional because it is justified by another agent, *B*.

1. Showing that a person has acted with intentions under another guise does not defeat *the GG* thesis.

2. A rational agent can still make a mistake. This mistake does not invalidate one's intention to pursue the worthy whenever one desires to act rationally.

3. People can intend/will/desire weakly because of other competing attractions not just from other possible guises, but also from *the GG*.

4. "Rational agent" as used generically for an actor of intentional action is a description of what intentional actors do.

Let us take a closer look at how compelling these responses are, and how they bear on our arguments thus far.

3. Objections from the GG

To raise the first of these rebuttals, it seems that merely showing that a person has acted intentionally under another guise does not defeat *the GG* thesis. Many reasons exist for why people act, and not all of them are admitted by the agents as good. Agents (as rational as they may be) can still admit that they are acting for specific bad reasons that make the said actions attractive to them. Milton's Satan comes to mind: an agent that is noteworthy for his contrariness, acting consistently on motives that are even acknowledged as bad by him. Whatever guise that can be identified to be opposed to *the GG*, it will still not have the magnetic feature that attracts actions that are worthwhile. In effect, what is done under any other guise can constitute anything but intentional action. There cannot exist any such intentional action that is not done sub *ratione boni.*[13] A weaker claim though goes thus: at least some representation of *de re* justification is necessarily required to act (Velleman, 1996). In this case, what is necessary to disprove the *GG* is not to show that some actions are done under other guises, but rather to show that an act can be intentional yet not be under *the GG*. On the face of it, this makes for a compelling narrative of intentional action. Indeed, unless there exists some justification or explanation for an agent's actions, it is hard to take that person seriously as an agent. We want to justify our actions not only to ourselves but also to others; and vice versa.

Secondly, mental and emotional states can impact rational agency. The attractions of other guises, second-order desires, pain, pragmatic considerations, and duties all can weaken an agent's motivation to act intentionally. As the Akan proverb goes: *Nsatea nyinaa nnyɛ pɛ*, (literally translates as *Not all fingers are the same*).[14] Not all attractions or justifications are the same, either. For instance, a statistician with erudition in proba-

---

13 Aquinas (*Summa* 1a2ae, 8,1).

14 The Akans are an ethnic group in West Africa with languages traceable to those spoken by people living in the ancient Ghana Empire around the 11th century.

bility rules will still gamble against their better judgment during a holiday in Las Vegas – while making experiments with the ten commandments. During these moments she does not lose all her mathematical prowess. She is merely pursuing pure pleasure in spending and having fun; no doubt these are strong attractions too. At worst she will be making mistakes, at best *the GG* is not in a profit and loss column on a spreadsheet. These attractions may not be justified. Be that as it may, they are still a force to reckon with. She may yield but she acknowledges she is acting against her better judgment. *Le matin je fais des projets, et le soir je fais des sottises* – in the morning I resolve, in the evening I commit follies. Yet this does not invalidate her intention to pursue *the good* whenever she desires to.

Third, there is the issue of *akrasia*, or weakness of will, which occurs when one acts against one's better knowledge. Now, it is not impossible that this should occur to rational agents (Davidson, 1970).[15] It is not impossible to see *akrasia* as rational, given some considerations. For instance, in the Biblical story of the patriarch Abraham (then Abram), we are told that upon arriving in Egypt to escape famine, he was asked about his wife by the Pharoah. Fearing for his people and his wife, Abraham merely feigned that she was his sister. [See (Genesis) 12:10-20]. Did he think lying was wrong? He must have, considering his acute sense of honor and uprightness, even reminding his God of the principles of morality on occasion (Genesis 18:16-33). Even a staunch Kantian will occasionally yield to a hypothetical imperative. In such moments, these actions are not exactly irrational to do. They were merely attracted by other desires they also value. This implies that the examples we have seen were the opportunity cost of choosing other competing desires. This co-opts the possibility of akrasia and sheds newer light on why one may yet be rational in moments of acting against what they know to be true and right. One can desire too much or too little.

Finally, the term "rational agent" is used as a description. The debate on action theory is rightfully centered on what real people do or purport to do. Reasons for acting, which we invoke in explaining *why* we act, are precisely what make actions intelligible. If these agents cannot rationalize to themselves or others any justification for what they do, they will strike us as very odd people indeed. When we talk about rational agents, we are not talking about any distinct biological class of entities. We are merely referring to the capacity of the human mind to make rational decisions. Given that this is a mere part of the whole of human mental activities, it is obvious why at least some actions by rational agents will not be rational. It is the mental states of these agents we speak of when we speak of rational agents (Audi, 2001).

Now, let us test these responses to our arguments. First, we will take a short detour

---

15 See also (Bratman, 1979); (Tenenbaum, 1999); (Arpaly, 2000); (Stroud, 2003).

through some examples and experiments.

SECTION 3

1. Experiments and Cases

(a) Salona is twenty (20). She wants to take heroin. As a medical student, she knows it may harm her and that she may develop an addiction. Nonetheless, she desires a hit.

(b) Roeber has been running since morning. John stops Roeber and asks why he is running. Roeber explains that he has received an injection from his doctor, who admits that the injection can cause people to be hyperactive.

(c) Nada went hiking with her colleagues. In a nearby cave, they get stuck when the entrance is sealed by a tremor. After unsuccessful rescue attempts, they decide to eat the flesh of whoever is selected by lottery. But they are all unwilling to kill since "Killing is morally wrong if one is not defending themselves against an immediate threat from an attacker". Like Buridan's Ass, Nada and eight friends die of starvation three days before their bodies are found by rescue workers.[16]

In (a), Salona is aware of the health implications of addiction. It turns out that she just enjoys the feeling of well-being and relaxation that narcotics induce. In (b), Roeber knows what he is doing yet cannot help but run. He does not have to, and he knows this. In (c), Nada does not act at all since she does not have justification. How do these exemplify the problems that I previously raised and the responses from the proponents of the *GG*?

2. Intentional actions that are not intentional?

Let us recall the argument for the thesis I am challenging:

P1   If action *x* is intentional, then agent *A* understands why she did *x*.

P2   If *A* understands why she did *x*, then *A* has justificatory reasons for doing *x*.

P3   If *A* has her justificatory reasons for doing *x*, then *A* sees *x* to be done under *GG*.

C   Therefore, if an action x is intentional, then *A* sees x to be done under *GG*.

I contend that following this reasoning an action can be both intentional and not intentional at the same time. Why? It is possible to understand an action without inten-

---

16   This is my adaptation of *The Case of the Speluncean Explorers* by (Fuller, 1949). The case is hypothetical, set in the year 4299: Roger Whetmore who was part of the trapped people in a cave is killed by his colleagues. This survival strategy seemed to have been suggested by the victim. Facing the moral dilemma of killing or dying, they chose to kill and survive. Rescued, they face trial and the philosophical implications of murder ensue in court. My adaptation is aimed at the opposite outcome: nobody gets killed because they believe killing to be wrong. As is expected they all die due to starvation. The aim is to assess whether in the practical world people hold on to their justificatory grounds to act, or when they lack justification they do not act at all like Buridan's Ass. Jean Buridan (c. 1301-1362) illustrated with a story how a starving ass placed precisely midway between hay and water simply dies of hunger and thirst because it does not decide which one to take.

tions about said action. Let us consider the case of Roeber running along the streets of South Bend. Roeber is aware of what he is doing. He even explains intelligibly to John why he is running. Did Roeber, from the look of things, want to be running? No. Given what we have discussed so far about intentions, it does not seem that an action that one does not deliberately initiate can constitute intentional action. Of course, we cannot infer from P1 that just because Roeber understands, he therefore intends – this would be a misinterpretation of the directionality of the conditional.

Yet while there may not be a relationship of causality between intentional action and understanding, I would suggest that they do share a relation of constitution. Suppose I said, 'If it rains, then the ground will be wet' and someone counters by saying "the ground is wet, but it did not rain outside", then they will be misinterpreting the directionality of the conditional statement. On the contrary, if I said: 'If you are a bachelor, then you are an unmarried man,' I do not mean being a bachelor causes one to be an unmarried man. Being an unmarried man is constituted in bachelorship. Similarly, the challenge I raised is pointing out the constitution of the consequent statement in the antecedent statement. If there is an asymmetry between the antecedent and the consequent statements (as we saw with the case of Roeber understanding but not intending an action), then we have a problem. Hence the challenge I raised has nothing to do with causation per se. What I desired to point out, like Diogenes, is that we have an unmarried man who is not a bachelor.[17] We have someone (like Roeber) who understands what they are doing and why without the action being intentional.

3. To act or not to act, that is the question

You do not need justification to act intentionally. In case (c) (i.e., under *Experiments and Cases in subsection 3.1*), Nada and her companions died because they thought some sort of justification was necessary in order for them to act.

Whenever rational agents are confronted with circumstances in which they lack justification to act, they do not always desist from acting. They may instead act without justification. In the crash case of the Uruguayan Airforce Flight 571 (which happened on October 13th, 1972), we find that survivors had decided to eat the deceased after the crash. Devout Catholics though the survivors were, they found nothing obstructing them from doing something that they believed to be totally unjustified.

One objection proponents *of the GG* could put forward is that the rational agents

---

17   If the rendition Diogenes Laertius gives us (in his Βίοι καὶ γνῶμαι τῶν ἐν φιλοσοφίᾳ εὐδοκιμησάντων) is to be believed, our edge-lord philosopher of Athens (also Diogenes) hearing Plato's definition of "man" as a "featherless animal" decided to pluck chicken and present it at *The Academy* saying: "Behold! A man." Plato corrected the definition by adding "...with broad flat nails." See in (Laertius, 1972), ISBN 978-0-674-99204-7.

in fact do have justifications for acting. They just do not acknowledge it or know about these justifications while they are acting. I call such justifications *ineffable justifications,* meaning that they are unthought of, unheard of, and unutterable. If such justifications exist, then in a way everyone who acts intentionally can offer her normative reasons, even as an afterthought. Post-hoc rationalizations of actions are not far-fetched to be invoked by rational agents, after all.

Response: This affords *the GG* no real solution. We remember from *Section 1* that actions are rational only insofar as agents understand why they did them. Moreover, justifications spring from the agent in question. Only their normative reasons, as understood by them, can justify their actions. This was an important point for *the GG* since it prevents us from making recourse to objective standards of justification. *The GG* thesis even stretches so far as to hold that "wrong facts" agents believe are proper justifications for their actions.[18] This includes the so-called "mistake of fact" such as pouring the wrong solution on your itching eye (Orsi, 2015: 714). So in order to maintain that rational agents always have agent-relative justifications for acting, *the GG* will have to posit that agents have *ineffable justifications* and that *ineffable justifications* are agent-relative.

But if the justificatory and explanatory reasons for acting depend on the agent, and if it is possible still for the agent not to know or have these justifications, then, what is the difference between a justification that is unthought of, unutterable, unheard of, and no justification at all? The so-called rational agents do not behave like Buridan's Ass and will act without justification. And if we say such justifications are hidden and unknown, *the GG* theorists also have to explain how agents come to the knowledge of them.

4. *Perverse Agents and the Good*

It is very easy to imagine people who act for specific bad reasons. Satan seems to act intentionally for reasons that are anything but good.[19] Some advocates of *the GG* attempt to explain such cases by suggesting that agents who act perversely act for the pleasure that is derived from acting for bad reasons.[20] But this is not precisely what is in question. I speak of the agent who admittedly acts for bad reasons just because they are bad.[21] Satan may enjoy acting for bad reasons, but could he derive a similar form of pleasure as a result

---

18   See (Parfit, 2001); (Alvarez, 2010).

19   There are also cases in which people act in ways that are neither good nor bad but just strange: Don Quixote, Anscombe's saucer of mud, Münchhausen's adventures, Voltaire's Pangloss. There is the kind of agent who can be flicking a light bulb out of mischievousness, or the kind who offers weak justifications for acting. Jack for instance [In (Golding, 1954), *Lord of the Flies*] says of his candidacy for chief: "...I can sing C sharp" (page 28). These are not problematic for the *GG* per se.

20   See (Tenenbaum, 2008); (Tappolet, 2009).

21   Raz (1999).

of acting for good reasons? It is not obvious that this is the case.

Moreover, it is not clear how simply understanding that one acts under the guise of the bad should cause one to act under the guise of *the good*. If acting intelligibly for bad reasons makes something an action under *the GG*, then this is a strange guise indeed. It is not uncommon to hear people say, "Yes, I know what I am doing is not right, but I want to do it anyway." We have learned that not only do people act contrary to what they perceive to be the right way to act in a given situation, but also that they do so deliberately. These are not exceptions to *the GG*. The difficulty that has confronted *the GG* has risen from common cases that jeopardize the necessary relation that proponents of *the GG* claim to exist between intentional action and the view that they are done for reasons considered good by the agent.

CONCLUSION

A separation of the justificatory and motivating or explanatory reasons for acting is sufficient to resolve most of the arguments against *the GG*. Much of this paper has centered on demonstrating an asymmetry between justificatory and the motivating reasons for acting. In addition to this, it appears that no specification of *the good* is given by *the GG* except the implication that agents are necessarily attracted to *the good*. But this good may be aesthetic, moral, or pragmatic good – or none at all. For instance, our example of Ponzi and Luther King rode on this leeway in *the GG*. Perhaps the difficulty for *the GG* is due to the complicated nature of its subject of inquiry: humans and their intentions. Perhaps suitable caveats will focus the debate on more specific aspects of the theory to dispel criticisms that may emerge only on tangential aspects.

In *Section 1*, I introduced the Guise of the Good (*the GG*) as an account of intentional action. *The GG* simply suggests that human beings as rational agents are pursuers of value: *the GG* explains intentional actions as those done necessarily for justifiable reasons. I highlighted that this necessary connection between intentional action and justification leaves *the GG* vulnerable to several attacks.

In *Section 2*, I demonstrated how people can act intentionally but for admittedly bad reasons. In this regard, separating motivating reasons from justificatory reasons is important: motivation is a matter of psychology, and justification is a matter of logic. With or without justification, intelligible actions can exist. No doubt justification is important, but it does not seem to be necessary.

In *Section 3*, we learned that if an agent is the final judge of the worthwhileness of her action, then *the GG* proponents cannot explain why the bad motivations for acting may become good. With the help of three experiments, I pointed out that human beings as rational agents may not always require justification to act intentionally. Advocates of *the GG* seem to not only acquiesce to the possibility of attraction from other guises but

also allow for "post-hoc" justifications. I challenged this response because there is no principled distinction between justifications the agent did not acknowledge and no justification at all. Thus, I suggest that *the GG* is not a convincing or necessary explanation of intentional action.

<div align="center">☙❧</div>

BIBLIOGRAPHY

Alvarez, M. (2010). *Kinds of Reasons*: An Essay on the Philosophy of Action. Oxford: Oxford University Press.

Anscombe, G. E. (1963). *Intention*. Oxford: Blackwell.

Aquinas, T. (1920). Summa Theologica. (F. o. Province, Trans.)

Aquinas, T. (2010). *Contra Gentilles*. Turnhout: Brepols Publishers.

Aristotle. (1925). *De Anima*. (W. D. Ross, Ed.) The Internet Classics Archive.

Arpaly, N. (2000). On Acting Rationally Against One's Best Judgment. *Ethics*, 110(3), 488-513.

Audi, R. (2001). *The Architecture of Reason*. Oxford: Oxford University press.

Ayer, A. (1936). *Language, Truth and Logic*. New York: Dover Publications, Inc.

Blackburn, S. (1984). *Spreading the Word*. New York: Oxford University Press.

Bratman, M. (1979). *Practical Reasoning and Weakness of the Will*. Nous, 13(2), 153-171.

Chambers, C. (2007). *Sex, Culture and Justice: The Limits of Choice*. Pennsylvania State University Press.

Dancy, J. (2000). *Practical Reality*. Oxford: Clarendon Press.

Davidson, D. (1970). Mental Events. In L. Foster, & J. W. Swanson, *Experience and Theory*. Humanities Press.

Fuller, L. L. (1949, Feb.). The Case of the Speluncean Explorers. *Harvard Law Review*, 62(4), 616-645.

King James Version. (1611).

Gibbard., A. (1990). *Wise Choices, Apt Feelings*. Cambridge: Havard University Press.

Golding, W. (1954). *Lord of the Flies*. New York: Perigee.

Hare, R. M. (1952). *The Language of Morals*. Oxford: Clarendon.

Hume, D. (1878). *Treatise on Human Nature*. (T. Green, T. Grose, & H, Eds.) London: Longmans Green and Co.

Kant, I. (2015). The Collected Works of Immanuel Kant: The Complete Works. PergamonMedia.

Korsgaard, C. M. (1986). Skepticism About Practical Reason. *Journal of Philosophy*(1), 5-25.

Korsgaard, C. M. (2009). *Self-Constitution: Agency, Identity, and Integrity*. Oxford University Press.

Laertius, D. (1972). *Lives of Eminent Philosophers* (Vol. 2). (R. D. Hicks, Ed.) Cambridge,

Massachusetts: Harvard University Press.

Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong*. New York: Penguin.

McDowell, J. (1985). *Values and Secondary Qualities*. (T. Honderich, Ed.) London: Routledge.

Moore, G. E. (1903). *Principia Ethica*. New York: Cambridge university Press.

Orsi, F. (2015). The Guise of the Good. *Philosophy Compass*, 714-724.

Orsi, F. (forthcoming). *The Guise of the Good. A Philosophical History*

Parfit, D. (2001). Rationality and Reasons. *Exploring Practical Philosophy: From Action to Values,* 19-39.

Plato. (1997). *Plato: Complete Works*. (J. Cooper, Ed.) Hacket.

Raz, J. (1999). *Engaging Reason: On the Theory of Value and Action*. Oxford: Oxford University Press.

Raz, J. (2011). *From Normativity to Responsibility*. Oxford: Oxford University Press.

Raz, J. (2016). The Guise of the Bad. *Journal of Ethics and Social Philosophy,* 10(3), 1-15.

Scanlon, T. (1998). *What We Owe to Each Other*. Harvard University Press.

Scanlon, T. (2014). *Being Realistic About Reasons*. Oxford University Press.

Scheffler, S. (2011). Valuing. In R. Jay Wallace, R. Kumar, & S. Freeman, *Reasons and Recognition: Essays in the Philosophy of T.M. Scanlon*. Oxford Scholarship online.

Setiya, K. (2010). Sympathy for the Devil. In S. Tenenbaum, *Desire, Practical Reason, and the Good*. Oxford: Oxford University Press.

Setiya, K. (2016). Akrasia and the Constitution of Agency. In Practical Knowledge: *Selected Essays*. Oxford University Press.

Shafer-Landau, R. (1998, July). Moral Judgments and Moral Motivation. *The Philosophical Quaterly* (1950-), 48(192), 353-358.

Shakespeare, W. (1992). *The Tragedy of Hamlet, Prince of Denmark*. (N. F. ed., Ed.) New York: Washington Square Press.

Smith, M. (1994). *The Moral Problem*. Oxford: Blackwell.

Staff, T. A. (2019, Nov. 7). *Study: It Pays to be Generous*. Retrieved from https://www.fool.com/the-ascent/research/study-it-pays-be-generous/

Stevenson, C. L. (1937). The Emotive Meaning of Ethical Terms. *Mind, 46*(181), 14-31.

Stocker, M. (1979). Desiring the Bad. *Journal of Philosophy*(76), 738-758.

Stroud, S. (2003). *Weakness of Will and Practical Irrationality*. (S. Stroud, & C. Tappolet, Eds.) Oxford: Clarendon Press.

Tappolet, C. (2009). Desires and practical Judgment in Action: Sergio Tenenbaum's Scholastic View. *Dialogue, 48*(2), 395.

Tenenbaum, S. (1999). The Judgment of a Weak Will. *Philosophy and Phenomenological Research, 59*(4), 875-911.

Tenenbaum, S. (2008, January). Appearing Good: A Reply to Schroeder. *Social Theory and Practice, 34*(1), 131-138.

Velleman, D. (1996). The Possibility of Practical Reason. *Ethics, 106*(4), 694-726.

ARCHÉ

A PROPOSAL TO THE STATESMAN'S AI REGULATION

by KYLIE FELICIANO

Boston University

ABSTRACT ‖ *Artificial Intelligence (AI) has become a pervasive force in our daily lives, from voice-activated assistants to large language models, the tools brought about by AI technology are quickly being adapted in companies and institutions throughout the world. This paper explores the urgent need for a robust regulatory framework for AI, grounded in ethical and philosophical insights. Building on Hans Jonas' 20th-century framework on ethical responsibility and incorporating perspectives from AI experts like Nick Bostrom and Mustafa Suleyman, this paper investigates the risks and challenges AI development at large poses to society. │ Bostrom's concept of "superintelligence" reveals the potential existential risks of AI, where its cognitive capabilities could surpass human understanding, leading to unpredictable and possibly dangerous outcomes. Suleyman, on the other hand, emphasizes the more immediate concerns surrounding AI, such as asymmetric impact, hyper-evolution, omni-use, and autonomy. Together, these perspectives underscore the need for ethical guidelines and safety measures that anticipate the rapid advancement of AI technology. │ This paper delves into Jonas' ethical framework and his concept of the "statesman's" or politician's responsibility, advocating for a forward-thinking approach to AI regulation. Jonas calls for a new gov-*

ernmental ethic that addresses the broader societal impacts of technology. The paper proposes a regulatory model that includes preemptive control in research labs, frequent audits, and collaboration between companies and governments to ensure the safe and ethical development of AI. The ultimate goal is to align AI development's trajectory with human values and societal welfare, establishing a global regulatory institution to oversee and coordinate these efforts. | Through this comprehensive framework, this paper seeks to mitigate the risks associated with AI while maximizing its benefits for humanity. The proposed approach involves rigorous oversight, international collaboration, and a commitment to responsible innovation, ensuring that AI serves the greater good and upholds the integrity of human life.

From iPhones that recognize our faces, Alexa's who understand our voices, and Chat-GPT at the helm of everyday life for many, AI is becoming an invisible yet indispensable companion in our daily routines. While we have seen enormous benefit in the tools that AI can provide individuals, companies, and institutions, beneath the face of convenience and efficiency lies immense ethical, social, and political complexities that demand urgent attention. The following paper endeavors to navigate some of these complexities, drawing upon the profound insights of leading thinkers in the field of AI and applying them to an ethical framework introduced by Hans Jonas in the 20th century. At the heart of this paper is an urgent call to action: to develop a robust and forward-thinking regulatory framework in the United States and beyond that not only keeps pace with AI's rapid advancement but also aligns its trajectory with the greater good of humanity.

In his seminal work "Superintelligence: Paths, Dangers, Strategies," Swedish philosopher Nick Bostrom warns about the profound implications of artificial intelligence as it moves towards the technological capability to surpass human cognitive capabilities across a spectrum of domains. Bostrom argues that upon the development of such technology, which he coins "superintelligence," human invention will become obsolete, as the superintelligent technology would vastly outperform humankind in every sector. With the onset of the epoch in which intellectual prowess of AI dwarfs the collective ingenuity of humankind, Bostrom posits that the greatest danger of superintelligent AI is the existential risk it could pose to humanity itself. According to Bostrom's Orthogonality Thesis, a superintelligent AI could potentially be directed towards any goal and there is no basis for an assumption that greater intelligence necessitates a greater capacity for anthropocentric ethical reasoning or benevolent behavior. An AI, unfettered by humanistic ethical constraints and propelled by indeterminable objectives, might employ its formidable cognitive mechanisms to achieve ends that are antithetical to human welfare. The profound capacity of such technology to strategize and adapt, could culminate in a

scenario in which its optimization process compromises human existence. Furthermore, Bostrom's analysis extends to sociopolitical ramifications in which the emergence of such superintelligence harbors the potential to create a power asymmetry, culminating in a single dominant superintelligent entity whose decision making and control over resources comes centralized and unassailable. In this grand scenario, Bostrom does not merely paint a picture of a dystopian world in which humankind is potentially subjugated by its own creation but incites a call to action for a collective effort in preemptively creating robust safety nets, encompassing AI goal alignment methodologies and the fortification of control mechanisms.

However, many academics believe thought about AI as an "existential threat" is merely distracting from the more immediate concerns. One such individual includes Mustafa Suleyman, British artificial intelligence researcher who is the co-founder and former head of DeepMind, an artificial intelligence company acquired by Google. In his book, *The Coming Wave*, Suleyman presents four features of AI technology which he believes pose more immediate challenges for governance, security, and societal structure-asymmetric impact, hyper-evolution, omni-use, and autonomy. Firstly, asymmetric impact, similar to Bostrom's concerns, highlights the democratization of power through technology on a more realistic scale. Suleyman asserts that such a democratization would enable individuals or small groups to exert influence at a scale historically reserved for nation-states. Additionally, the immediacy of such an effect is exacerbated by the potential for asymmetrical warfare and the disruption of conventional deterrence models. For example, AI-driven cyber warfare could disrupt critical infrastructure and undermine traditional national defense mechanisms, which previously relied on deterrent effects of mutual assured destruction or economic and technological superiority. Secondly, hyper-evolution, which refers to the accelerated pace at which AI technologies develop and improve, encapsulates not only rapid advancements in computational capabilities and algorithms efficiency, but also the integration of AI into a wide range of fields and applications. With accelerated AI adoption, legislative and ethical guidelines may not be able to keep pace with such advancements, drastically altering and leaving vulnerable job markets and economic structures. Similarly, AI's omni-use, or the general-purpose nature of AI, can render itself both beneficial and detrimental if a granular approach to regulation is not created. For instance, AI algorithms designed for data analysis can be used in healthcare for predictive diagnostics while simultaneously being used for surveillance with privacy implications. This poses a large challenge for regulation as the same AI tools can be used for vastly different purposes, good and bad. Lastly, autonomy in AI systems raises profound questions about agency, responsibility, and the feasibility of human oversight. Autonomous systems are able to interact with their surroundings and take actions without the immediate approval of humans. New forms of autonomy have the potential to produce a set of novel effects, not only unfamiliar to its creators, but

impossible to trace backwards. Previously, creators could explain how something worked and why; however, this is increasingly no longer true, further exacerbating the problem of regulation as governments may find it difficult to regulate technologies that operate beyond human oversight. Synthesizing the ideas from Bostrom and Suleyman, it is clear that we have a right to be concerned. There is a strong case that an AI superintelligence would be fully impossible to control or constrain and the blunt truth is that nobody knows when, if, or how this may happen. There very well may come a point where technology can fully direct its own evolution; where it is subject to recursive processes of improvement; where it passes beyond explanation; where it is consequently impossible to predict how it will behave in the wild; where, in short, we reach the limits of human agency and control. With this in mind, it is crucial that we pour into regulatory work, not only to govern current applications, but also to anticipate and shape future trajectories of AI development. The convergence of Bostrom's existential warnings and Suleyman's immediate practicalities calls for an integrative approach to regulation—one that accommodates the dynamic and unpredictable nature of AI, ensures the alignment of AI with human values, and manages the risks associated with AI's rapidly advancing capabilities. In the following sections of the paper, I will explore the philosophical backing behind why and how ethical regulation should occur for the governments in which such technologies are developed using the lens of Hans Jonas' The Imperative of Responsibility.

In Hans Jonas' 1979 *The Imperative of Responsibility*, Jonas urges humanity to reevaluate its inherent responsibilities in a new age of technology and more importantly, a new age of ethics. In the post-WW2 era, in which Jonas had served for Britain's Jewish Brigade in Palestine to fight against Hitlerism, he reflected on the new warfare technologies he had witnessed in his early life. Jonas begins with a review of previous traditional ethics which he says granted the non-human world to humankind to manipulate at their whim. Dealings with nature were ethically neutral and it is the entity of the human that is the constant and could not be reshaped by *teche*. Further, doing good or evil deeds are limited to the action in itself, which means it is limited in time and space. He then makes the claim that a technological revolution has occurred in which the fate of humankind can now be manipulated by techne itself. It is because of this revolution that we must instill a new ethic into our everyday lives. No longer are the days of "neighbor" ethics all we must concern ourselves with, but now, because of the newfound vulnerability of humankind, we must begin to consider cumulative ethics as well, or rather one that considers the future. However, Jonas stumbles upon the question of who is responsible for such technology which he answers through the use of an analogy of a parent and a statesman to illuminate the innate responsibility humanity has in its interactions with any invented technology, drawing parallels between the care of a parent for their child and duty of a statement to safeguard the welfare of society and future generations.

Jonas delves into the theory of responsibility to form the basis of distinguishing between the responsibilities of the individual (the inventor) and the government. Firstly, he distinguishes between formal responsibility, which is about being accountable for one's actions and their consequences, and substantive responsibility, which concerns the positive duty to act and care for others. For natural responsibilities, the immanent "ought-to-be" of the object claims its agent a priori; an example of this is parenthood. For the substantive responsibility, the responsibility is a posteriori, given that the terms of the relationship are entered into. In moral status, the natural responsibility is stronger as it is the original from which any other responsibility derives its validity. When it comes to substantive responsibilities that have been instituted "artificially" by bestowal and acceptance of a task such as an appointment to an office, its acceptance has an element of choice that does not exist under the parental example. Important in the latter is that responsibility draws its binding force from the agreement that created it and not from the intrinsic validity of the cause. Jonas says that for responsibilities that are merely stipulated and not dictated by an intrinsic claim of the cause, behavior contrary or negligent of the stipulation is not necessarily inherently irresponsible. Jonas reserves the word irresponsible for a betrayal of responsibilities of the independently valid type by which a true good is endangered... However, Jonas argues that even in the case of a stipulated responsibility, such as a tax official, the general ought-to-be thesis holds as the cause is primary in responsibility or in other words, insofar as the ultimate object of his responsibility, beyond the direct one, thus is the true cause of the upholding of loyalty relations in the general upon which society rests. Jonas says this social contract is substantive, inherently obligating good. Thus, the negligent official who is in violation of duty becomes irresponsible after all.

However, Jonas argues that there is a third responsibility that is unique to human freedom and goes beyond the inherent and contractual responsibilities of natural and substantive responsibilities. In the case of a statesman or politician, the responsibilities that come with such a position are not chosen, but rather they are inherently part of the role due to the effective range of the politician's power. This self-chosen responsibility is when a person identifies a moral or ethical good that is not immediately within their power to effect and then chooses to take on the responsibility for it anyway. For Jonas, this is a good in the first order or rather a fundamental good that carries the inherent "natural" rightness about it. Most importantly, Jonas asserted that the parent and the statement are eminent paradigms in that the primary objects of responsibility are other human beings which entails an inherent value placed on human beings' existence with respect to continuity and a future. While parents are responsible for the direct care and nurturing of their offspring, with a clear beginning and end to the duty – specifically, when the child reaches adulthood, statesmen have a responsibility that extends beyond direct relationships and reaches collective needs. These decisions and actions

have far-reaching consequences that affect not only the current collective, but future ones as well. This relationship does not have a predefined end as although it persists for the period they hold power, the effects continue through their decisions. Thus, in combination with the rise of devastating technologies, Jonas calls for a new governmental ethic of foresight and care for the future, which takes into account the potential impacts on human life. Additionally, Jonas discusses the intrinsic responsibility of humans, in that by virtue of being causative agents, inherently have a sphere of influence and thus carry an obligation that precedes moral or immoral action. Jonas says to be de facto responsible in some respect for someone at some time (whether acknowledging it or not) belongs as inseparably to the being of humankind as their a priori capacity for it. In this sense an "ought" is concretely given with the very existence of humankind, as being a causative subject involves itself objective obligation in the form of external responsibility. With this responsibility, they are not yet moral, but instead members of the moral order. Thus, applied to the innovator or inventor there must be a visionary commitment to a good that may not yet exist within the current capacity or societal framework. In the remaining sections of this paper, I will illustrate what a practical model of this ethical framework could look like in the context of AI regulation, specifically within the duties of the statesman, as individual action is near impossible to ensure or control.

As articulated previously, the need for a regulatory system for AI is imminent, as the technology is threatening both the immediate fabrics of our society and in more dystopian scenarios, human existence itself. In Jonas' philosophical framework, the statesman is entrusted with a prescient responsibility that anticipates the long-term implications of policy decisions, particularly those that bear upon the ethical and societal dimensions of technological innovation. This forward-thinking stewardship obligates the government to enact and enforce regulatory measures that align AI development with human values and societal needs, mitigating risks and maximizing benefits. The unique challenge presented by omni-use hyper evolving technology necessitates a foundational shift in inception. Thus, our regulation and goaled containment of AI technology must start in the lab. To prevent bad actors (the inventors) from disregarding their inherent responsibility, ethical regulatory frameworks for research must be established in the same way they exist for biomedical research. To begin, we must place a focus on developing technologies that are narrowly scoped rather than omni-use. For example, a nuclear weapon is highly specific technology with one purpose, thus making it easier to contain. For those AI technologies that can be applied towards incredibly general purposes, we must discourage and reconsider development, which would need to be approved by a larger governmental board as is with other sectors of scientific research, to aid in the prevention of market disruption. In terms of the problem of hyper-evolution, in which inventors no longer have the capability to understand their inventions, there must be formal mechanisms within research institutions to help determine the extent in which

a technology can be dematerialized. For this, research must begin to evaluate the permissible bounds of understanding, development past the realm of understanding should be put into serious question. Further, researchers must consider the necessity of the AI solutions in question. If viable alternatives to the problems they are trying to solve with AI technology are already in existence they should not continue to be developed. For example, chlorofluorocarbons or CFCs are non-toxic, non-flammable chemicals containing atoms of carbon, chlorine, and fluorine that were once debated to be put in use as refrigeration methods. However, they were banned partially because there were cheaper and safer alternatives for refrigeration. The same must be applied to AI technologies. Additionally, a system to measure asymmetric impact and the autonomy level of the technology should be developed. If the technology has a certain number of autonomous characteristics or, in other words, the scope for self-learning and operation without oversight is too large, the development should be reconsidered and most likely denied. The more the technology by design requires human intervention, the less there is a chance of losing control over it. Lastly, it should also be considered if there are sufficient resource or engineering constraints on the technology's invention, development, and deployment in a border sense. The availability of resources to fuel the advancement of AI bears directly on the potential for near-term containment and control. Overall, the proposed regulatory paradigm would be enacted in AI research, extending to formal academic labs and private companies prior to the development of such technologies and would require that such development be both conscientious and constrained.

Going further, technical safety advances and regulation as proposed must be frequently verified as being effective, which can be done through audits. Currently, there is no formal or routine effort to test the systems of development that exist. Suleyman says, "As a starting point, having companies and researchers working at the cutting edge, where there is a real risk of harm, proactively collaborating with trusted experts in government-led audits of their work, is basic common sense" (Suleyman, Coming Wave, 305). These verification systems would be analogous to the routine performance audits in other high-stakes industries like the pharmaceutical industry, where cost of failure is human life. Proactive collaboration between companies, academics, and regulatory bodies is absolutely necessary for transparency in AI development. These actors should be legally bound to share knowledge of emerging risks, similar to how cybersecurity threats are handled, in order to foster a culture of safety. Aside from this collaboration, a more tangible addition to this idea is "red teaming" or groups specifically hunting for flaws in AI models and software. Government-funded red teams would ensure that insights discovered along the way are shared widely across the industry. Eventually, this work could be scaled and automated, with publicly mandated AI systems designed specifically to audit and spot problems in others, while also allowing themselves to be audited. Furthermore, the implementation of secure, real time reporting systems offers

a promising avenue for maintaining oversight. SecureDNA is a non-for-profit program started by a group of scientists and security specialists whose goal is to plug every DNA synthesizer—benchtop at home or large and remote—into a centralized, secure, and encrypted system that can scan for dangerous pathogenic sequences. If people are printing potentially harmful sequences, they're flagged. The program is cloud based, free, cryptographically secure, and updates in real time. The establishment of an AI technology program modeling this one would be a vital tool in preemptively identifying and mitigating risks, especially as AI capabilities continue to advance so widespread and rapidly.

The basis of the previously proposed ideas is that these systems of regulation, auditing, and actor cooperation would be written into law, and for this reason it is crucial for the state to maintain a strong capability to do so. This strong capability means extending the role of government beyond passive oversight and necessitates active participation in the formulation and maintenance of ethical standards. The state must not only legislate, but actively participate. This means investing in the technical acumen needed to fully understand and keep up with the intricacies of AI development, embodying Jonas' ideal of the statesman who wields the proper knowledge and authority to shape his decisions prudently. The commitment to this cause must be substantiated by adequate funding, reflecting a recognition of the far-reaching impact that AI has on every facet of society, economic to social. Such funding will allow governments to hire in-house expertise as well as incentivize responsible innovation through grants and funding programs for research that aligns with societal welfare. Overall, with a comprehensive understanding of just how such technologies work and are developed, adequate legal frameworks can be developed with more ease and a global standard can begin to be set.

To this end, there is a forward-thinking proposal on the table to collaborate internationally and create global institutions that would be singularly dedicated to overseeing the realm of such technology. These bodies would parallel the foundational principles of the World Bank or the United Nations, serving as policy architects and international regulators for the ethical development and deployment of AI in various industries. Such a body would serve as a catalyst for global collaboration, setting standards and fostering a world where technology and human interests can come together, allowing for a culture of sustainable and positive innovation. Ultimately, it is the responsibility of Jonas' statesman, the government, to not only inform itself but be proactive in order to enable a deep commitment to ethical deliberation, ensuring that capabilities of AI are harnessed for the collective good, preserving the integrity of human life and the societal fabric that sustains it.

☙❧

BIBLIOGRAPHY

Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2014.

Jonas, Hans. *The Imperative of Responsibility: In Search of An Ethics for the Technological Age* Chicago: University of Chicago Press, 1984.

Suleyman, Mustafa. *The Coming Wave*. Crown Publishing Group, 2023.

ON THE EPISTEMOLOGICAL SIGNIFICANCE OF BODY NARRATIVES

by Santiago Flórez Sánchez

Universidad de los Andes

ABSTRACT ‖ *This paper explores the relationship between identity, narrative, and the body within oppressive contexts. I argue that various identity groups—defined by gender, ethnicity, and ability—are epistemologically marginalized due to judgments based on their physical bodies. By distinguishing the body as both a physical object and an experiential subject, I contend that its social-epistemological significance relies on the identity and narrative it embodies. Finally, I examine the ethical dimensions of body modifications and emphasize the profound impact of body narratives on social imagination.*

The body is to be compared, not to a physical object, but rather to a work of art… [Works of art] are individuals, that is, beings in which the expression is indistinguishable from the thing expressed, their meaning… It is in this sense that our body is comparable to a work of art. It is a nexus of living meaning…[1]

I. INTRODUCTION

Although some identities are not founded on body configurations, the body is related (albeit not always directly) to several identity groups. Take, for example, identities based on sexual orientation, gender, ethnicity, and ability.[2] These identity groups are related to the body because this is the substrate of the agent and its individuality. It is worthwhile to deepen our understanding of oppression and marginalization[3] based on the body. An oppressive context on account of identity marginalization supposes the distinction and exclusion of an identity group, which, in turn, often leads to the distinction and exclusion of certain bodily features, since the body is the most noticeable and underlying point of comparison. This is clearly the case when a woman is harassed or a disabled person is overlooked, but it is also involved when a Black, Indigenous, trans, female, or disabled person is epistemically underrated. Accordingly, this essay focuses on the distinction between the body as a physical object and the body as an experiential subject and aims to explore the socio-epistemological consequences of those bodies that express marginalized identities – namely resistant bodies.

In the first section, I will show that judgments about one's knowledge based on one's physical capacities are typically unfounded, even though it is accepted that there is a causal link between physical and mental states. In the second section, I am going to connect the body with experience and experience with narrative, thereby presenting the relationship between body and narrative. Here a narrative is understood as something that tells a story, that allows us to find out things about a person.[4] In this sense, the natural transformation or artificial alteration of the body may contain a narrative. The idea is that experiential bodies determine and express narratives while also being narratives themselves (i.e., body narratives[5]). This discussion will lead me to question to what extent body modifications are morally acceptable. Finally, the conclusion will explore how body narratives affect the social imagination. The importance of this inquiry is premised on the fact that, in a basic sense, identity marginalization relies on the body and, consequently, any discussion about how to make a more inclusive society has to take into account body narratives.

---

1 Merleau-Ponty, *Phenomenology of Perception*, Pt. I, Ch. 4.

2 The clearest case is in regards to disability, insofar as it is often defined in terms of body configurations and its relationship with the world (Garland-Thomson, 2012, pp. 341-342).

3 Here I understand oppression as an unfair power relationship between people, regardless of whether it is individual or structural. In turn, marginalization is a mechanism of oppression that aims to segregate oppressed people.

4 Anzaldúa, "now let us shift … the path of conocimiento … inner works, public acts;" Pitts, "Gloria E. Anzaldúa's Autohistoria-teoría as an Epistemology of Self-Knowledge/Ignorance."

5 It is important to remark that, in this paper, body narratives are not understood as narratives about bodies, but as bodies being narrative.

II. THE BODY IN OPPRESSIVE CONTEXTS

I am interested in arguing why judgments about intelligence based on the body as a physical object are usually[6] unjustifiable. I will state that some identities are marginalized because the oppressor denies the person (and their experiences) beyond a physical body. This leads to an improperly established link between judgments relative to the physical body—like being Black, female, or disabled— and judgments about intellectual capacities and cognitive traits. In the following, I shall pose racist, sexist, and ableist cases to illustrate these ideas.

To begin, it should be noted that, during slavery and colonization, Black and indigenous people were oppressed by people with "radically different" bodies. Different bodies likely helped enslavers and settlers refrain from empathizing by focusing on Black and indigenous bodies as mere physical bodies or usable objects.[7] This brings me to the distinction between the body as a physical object and the body as an experiential subject. It is not an ontological dualist distinction, but a distinction between the third-person and the first-person perspectives. Therefore, these are not exclusive notions: people perceive their own and others' bodies as material objects, hopefully without denying (as done by enslavers and settlers) that others also have a meaningful first-person experience. According to Bakare-Yusuf (1999), the slaver's logic relies on the brutally violent transformation of the liberated body to the captive body, and this transformation entails a distinction between "the body of knowledge and the body of labour"[8]:

> [In the U.S.] The slavers did not intend the slaves to die; they were concerned to ensure the survival of the black body —albeit in a demoralized form. The black female body is a useful body because it is both a labouring, sexual and reproducing body and therefore it was necessary to preserve the health of the enslaved woman. The use of violence was therefore necessary to break them in, to fragment them, to destabilize them and to *make them cease to be subjects, to transform them into 'docile bodies'* (Foucault 1977) that became bodies that labour.[9]

The lack of subjectivity of the colonized and enslaved body cuts off any possibility of

---

6 Even if such judgments are commonly unfounded, a limitation of this argument (regarding some neuroatypical difficult cases) will be considered below. That is why the conclusion cannot be universal.

7 Interestingly, a similar interpretation is compelling in regards to nonhuman animal treatment. This would be the specist case: humans see other animals as objects (i.e., physical bodies) they can use, and they avoid empathy and normalize the poor treatment of nonhuman animals because they have radically different bodies. Nonetheless, this paper is restricted to human groups.

8 Bakare-Yusuf, "The Economy of Violence: Black Bodies and the Unspeakable Terror," 311.

9 Ibid., 318, emphasis added.

knowledge: objects do not and cannot know anything. To the extent that Black bodies were constructed as physical objects (i.e., bodies of labor), Black people were excluded from the possibility of producing and acquiring knowledge. In other words, their experiences and intellectual capacities were invalidated. The same can be said about the female body since sexism reduces it to a reproductive and/or sexually desirable object. Once again, it is the denial of a meaningful, experiential body (i.e., a subject) which allows oppressors (e.g., slavers, settlers, sexists, ableists) to block empathy[10] and to be violent. In short, identity marginalization is founded on the oppressor's perceived distinction between bodies that classify as subjects (i.e., experiential bodies) and mere usable objects (i.e., physical bodies). This is the foundation of the epistemic injustices proper of such oppressive contexts, insofar as only subjects are considered epistemic agents.

Of course, the correlation between marginalized bodies and inferior cognitive capacities is illusory. To try to justify a cognitive inferiority of marginalized bodies, oppressors presuppose the same bias: that such marginalized bodies cannot bear meaningful experiences. Okruhlik (2013), for example, speaks to research on gender differences in intelligence:

> I have in mind here recent developments in neuroanatomy which are directed to explaining intelligence differences between women and men [...]. [...] What is interesting for our purposes is that for many researchers the one element of the theoretical network they are unwilling to surrender in the face of recalcitrant data is the assumption that there must be predominantly *biological* reasons for inferior intellectual achievement in women.[11]

Okruhlik argues that conceiving women as cognitively inferior to men precedes objective reasoning because such a conception is the response to one's incapability of seeing beyond the physical body. The oppressor's belief that the oppressed is cognitively inferior is the result of a careless omission of other people's capacities and experiences. It is an inconsistent judgment on the part of the oppressor that endorses their ensuing violence of marginalization. The tendency to judge others as epistemically inferior, by virtue of their body configuration, is proper in oppressive contexts, but it does not have

---

10 Some have argued that empathy could be detrimental in regards to disabled people. According to Garland-Thomson (2012), "Because empathy depends upon the experiences and imagination of the empathizer in regarding another person, prejudices, limited understandings, and narrow experience can lead one person to project oversimplified or inaccurate assessments of life quality or suffering onto another person" (p. 350). This idea is compatible with the present point of the argument, since the oppressor's focus on the physical body refers to a previous state in which empathy is absolutely neglected, but it does not entail that empathy alone would eradicate oppression and violence.

11 Okruhlik, "Gender and the Biological Sciences," 26-27.

any real justification.

It is possible to expand the analysis from bodily intrinsic factors to bodily normative factors. Oppressors marginalize groups by reducing them to their physical bodies and assume marginalized groups are epistemically inferior because of their very bodies. Nonetheless, there are cases where oppression is not directly based on the body, but on the presentation or performance of the body. While being Black or female is related to observable bodily features (i.e., bodily intrinsic factors), following certain manners or dressing in a certain way are bodily normative factors.

Consider, for example, the transphobic violence broached by Professor Talia Mae Bettcher (2014), drawing from her own experience as a trans woman. "Reality enforcement" refers to the identity invalidation of trans people which is based on the presumed appearance-reality contrast. In other words, the term "reality enforcement" refers to the violent conception of trans people as deceivers, and it is founded on a bodily normative factor: the presupposed representational relationship between the public gender presentation and the intimate self-presentation (i.e., genitalia).[12] Due to this violent conception, trans identities and experiences are invalidated by interpreting their body presentations and/or performances as misleading others about their "real" genitalia. This normative interpretation invalidates the identities of trans people because it assumes that their body presentations seek to inform others about their genitalia. According to Bettcher, such a representational relationship "is part of a larger system of nonverbal, non-consensual communication that facilitates (hetero)sexual manipulation". [13]

Oppressors also marginalize groups by conceiving them as breaching the norms of body presentation and performance. They then assume marginalized groups are epistemically inferior because of their body presentations or performances. The ascribed cognitive inferiority of trans people by a dominant group is evident in the rejection of trans knowledge —for instance, the idea that gender is a passable space— and in the frequent classification of transsexuality as a pathology. Thus, marginalized identities and the knowledge of marginalized groups are judged and excluded not only because of what their bodies are but also because of how they present or move their bodies (depending on the normativity of the oppressive dynamic).

Oppressive contexts establish a normative framework dictating which bodies, presentations, and behaviors are deemed acceptable for epistemic agents, whose experiences and knowledge are considered valuable. Bodily intrinsic and bodily normative factors on which oppression relies overlap in most cases, but each factor implies different conditions. While a sexist context condemns a transwoman for deviating from traditional

---

12 Bettcher, "Trapped in the Wrong Theory: Rethinking Trans Oppression and Resistance," 391-393.

13 Ibid., 394-395.

gender norms, it also condemns a cisgender woman for having certain genitalia. In the same way, both lead to baseless judgments about the epistemic capacities of certain bodies.

To clarify things, the case of (dis)abled people may help. Bioethicist Jackie Leach Scully (2018) argues that "disabled people often comment that others 'don't look past the disability.'"[14] This is an example of the same kind of denial of the subject's inability to see beyond the physical body, as aforementioned. Such a denial involves an illogical ascription referred to by Scully as the "global epistemic incapacity."[15] Scully's global epistemic incapacity refers to the assumption that a disabled person is incapable of independent thought or decision-making due to a specific physical disability. In this sense, it is the same kind of judgment that moves from a body configuration or performance to an assumption about that body's cognitive capacities. The physical body is related to both social marginalization and epistemic exclusion in cases of disability, gender, and ethnicity.[16]

Let us consider a possible counterargument. Since there is general agreement about the supervenience of mental states on physical states, judgments about mental capacities based on the body may not be irrational. Nonetheless, specific isolated bodily features relevant to the marginalization of people in oppressive contexts are *typically* unrelated to (and thus cannot justify judgments about) cognitive capacities. It is necessary to introduce two nuances. First, Okruhlik's case did not intend to imply the impossibility of establishing certain body configurations as cognitively more capable than others. The real problem flagged by Okruhlik was that in oppressive contexts the reasoning behind the ascription of cognitive inferiority is often lacking, not that a well-founded ascription of cognitive inferiority to certain body configurations is impossible. Surely, the question of whether body configurations are related to cognitive capacities could be treated as an empirical scientific investigation. Second, even the defending thesis —that judgments about physical features relevant to social marginalization cannot justify judgments about cognitive capacities— has its limits with respect to some mental disabilities. In Scully's own words:

> The complicating factor in this argument is that some learning disabilities, severe developmental disorders, and mental health problems genuinely do compromise a person's ability to comprehend their world (or, perhaps, more accurately, their ability

---

14 Scully, "From 'She Would Say That, Wouldn't She?' to 'Does She Take Sugar?' Epistemic Injustice and Disability," 116.

15 Ibid.

16 An interesting counterexample is the case of Asian Americans, who face positive stereotypes regarding their cognitive abilities, particularly in STEM fields. However, this stereotypical assumption is also unfounded, usually considered abnormal, and does not preclude other negative judgments about cognitive features.

to comprehend the world and communicate it in the same way as the majority).[17]

III. THE BODY TRAVERSED BY EXPERIENCES

We explored the body as a physical object and its role in judgments about cognitive capacities. We will now turn to look at the body as an experiential subject. The epistemological significance of a body is not related to its materiality but to its experiences.

This section is divided into three parts. First, I will consider the link between the body and experience. On the one hand, experience is typically not confined to the present, but extended to the memories of one's past and the possibilities of one's future. On the other hand, the body determines the frame of one's possibilities. Second, I will examine how the body's experience is embedded in narratives. I will discuss three points: (1) the transformation of the body's experience through narrative, (2) the meaning-making processes implied by narrativity, and (3) the opening of narratives to the collective and narratives' subsequent scrutiny. Even though a body's experiences and narratives share a temporal structure, they differ with respect to their private and public characters. The body's experience is subjective, meaning that no one can experience the specific first-person perspective of another. In contrast, narratives are objective because they imply a reflective (in this sense, third-person) reconstruction of experience. Finally, I will delve into the embodiment of narratives. What is a body narrative? What is the role of body modifications in regard to body narratives? To deal with these questions, I will examine the case of tattooed bodies. I will make clear that tattoos are not mere pictures or words arbitrarily located on the body. I will then explore to what extent body modifications are acceptable. To answer this question, I will explore sex-change surgery and healthy limb amputation.

1. Experiences inside the Body

The denial of the subject —not recognizing an experiential body beyond the physical body— was considered the foundation of marginalization based on the body. Compared to the physical body, the experiential body is deemed to be proper for subjects. The experiences we have and the meanings we construct are necessary for personhood. It would be nonsense to ascribe subjectivity or first-person experiential knowledge to inanimate objects. Marginalized bodies are judged as cognitively inferior since they are conceived as lacking meaningful first-person experiences. The marginalized body, as perceived by oppressors, does not convey the intricate complexity of meanings that a subject experiences. However, as subjects, every conscious body is always "looking" for meaning

---

17 Scully, "From 'She Would Say That, Wouldn't She?' to 'Does She Take Sugar?' Epistemic Injustice and Disability," 117.

through their experiences, and there is a constant and usually unconscious effort to maintain or recover meaning. This meaning-making process (previous to language)[18] is constitutive of experience.

Meaning in experience is temporally constructed. Time consciousness is the basis of subjectivity. Husserl (1973) argues, "When we inquired about the connection which makes possible the unity between all the perceptions and positional presentifications of an ego, this was found to be the *temporal connection*."[19] A multiplicity of experiences is arranged according to the temporal structure in order to compose an experiential body (subject). Each experience is organized according to its temporal relation. All of the body's experiences are part of the same subjectivity despite the fact that they are differently apprehended: one as present, another as recalled, another as projected, and so on. Thus, time consciousness and meaningful experiences rest on two crucial functions of awareness: retention and projection. Subjects experience the present as marked by the memories of preceding experiences (retention) and by the expectations of the experience's unfolding (projection). In short, meaning depends on relating and organizing experiences into a temporal structure and, in turn, relies on bearing in mind the previous experience (producing a familiarity) and anticipating the following experience (in relation to a frame of possibilities).

It is important to note the link between experience and activity. To illustrate, consider ordinary experiences at home. One's experience is molded by a frame of perceptual and performative possibilities. One enters the kitchen with an expectation for how the experience is going to be, according to one's familiarity with the furniture's style, position, and so on. In addition, entering the kitchen presents several possibilities for action, like grasping certain objects, moving through certain paths, cooking, eating, cleaning, etc. Insofar as one's actions are performed by means of one's body, one's particular body configuration makes possible some particular actions, and the same could be said about perceptual possibilities. As Garland-Thomson suggests, "our bodily form, function, comportment, perceptual apprehension, and way of mind shape how we understand our world."[20] In other words, the configuration of the body shapes one's whole experience by defining the possible actions and subsequent experiences. In this sense, the body is the basic ground of experience, and experience is private in that it is directly correlated to a specific body.

---

18 The level in which the phenomenological tradition (Husserl, 1973; Merleau-Ponty, 1962) describes conscious experience is much more basic than the linguistic aspects of experience. In this sense, it may refer to a perceptual experience similar to one of a nonhuman animal.

19 Husserl, *Experience and Judgment*, §38.

20 Garland-Thomson, "The Case for Conserving Disability," 345.

2. The Expression of Experiences

After connecting the body with experience, it is necessary to relate experience with narrative. While various types of narratives (such as fictional, theoretical, poetic, personal, etc.) can be differentiated, I will focus on the type most closely connected to first-person experience. Drawing on Anzaldúa's (2009) idea of *autohistoria-teoría*, Pitts (2016) refers to self-writing as an exercise of analyzing one's own personal experiences and constructing self-knowledge. Pitts acknowledges three key features of self-writing: (1) the transformation of personal experience into narratives (the sensuously embodied aspect), (2) the meaning-making processes implied by narratives (the productive aspect), and (3) the opening of narratives to the collective and its scrutiny (the collaborative aspect). These key features of self-writing overlap in various ways.

The idea that "the generative work of narrative is to produce knowledge through rendering life experience into coherent and usable form" means that the transformation of personal experience into narrative is central to epistemological concerns.[21] Narratives rest on experience insofar as they transform experience by meaning-making processes into a "coherent and usable form," that is, as knowledge. This is the sensuously embodied aspect of self-writing and reflects the fundamental relationship between personal experience and narrative in general. According to Pitts, "[epistemic situatedness] depicts knowledge practices, including self-knowledge practices, as dependent on how we exist in the world as concretely embodied beings."[22] Pitts, however, does not exclude fictional, theoretical, or poetic types of narrative. After all, since narratives are built on personal experience, they also include aspects of emotion, imagination, conceptualization, perception, and action.

The meaning-making processes implied by narratives highlight the productive aspect of narratives. Experience itself is a meaning-making process in a radically different sense, as suggested earlier. The constant looking for meaning is a surreptitious and (most of the time) unconscious process, whereas the creation of a narrative is a conscious and hopefully reflective process of meaning-making. In the former case, the search for meanings entails, for instance, the constitution of an object by drawing on its diverse presentations, whereas, in the latter case, the creation of a narrative gives meaning to complex experiences, where different entities and happenings are interrelated. Narratives rely on the experiential level and language in particular, as only in language can such a complexity be represented. This indicates the public character of narratives.

The collaborative aspect of narrative is associated with its public character. Since language is essentially public,[23] narratives, which depend on language, open the possibility of complex communication with others. Pitts claims that "writing about oneself provides the theoretical tools for others to critically interrogate their positions and the world."[24] Consider the famous example of the coining of the term "sexual harassment."[25] This is a case in which storytelling leads to filling "a gap in the collective hermeneutical resource."[26] Such a positive consequence of narrativity is not necessary, but the importance of the collaborative aspect remains: a narrative is introduced by its nature into the collective, so that it may make others reflect their narratives. Anzaldúa summarizes the three points —from her discussion about autohistoria-teoría— as follows:

> By writing about the always-in-progress, transformational processes and the constant, ongoing reconstruction of the way you view your world, you name and ritualize the moments/processes of transition, inserting them into the collective fabric, bringing into play personal history and fashioning a story greater than yourself.[27]

3. The Body Made Narrative

I have described how body configurations impact personal experience and how narratives are founded on one's personal experience. But how could a *body be* a narrative? To understand the embodiment of narratives, it is useful to take into account the dependency of narratives on meaning-making processes. This may help distinguish body configurations that express narratives from those that do not.

A physical object cannot be a narrative. Even though physical objects are open to the collective, they have no link with personal experience and thus do not involve any meaning-making process by themselves. If physical objects are not narratives, then are experiential subjects automatically narratives? Pre-linguistic humans and nonhuman animals are experiential subjects since they have meaningful experiences; nonetheless, the meaning-making processes in which they are involved do not reach the characteristic complexity of a narrative because they are pre-linguistic. I argue that a distinction is needed between expressing a narrative and being part of a narrative. Pre-linguistic humans and nonhuman animals can be part of a narrative if they are interpreted by others, but those narratives to which they are related are not their narratives because

---

21 Ibid.

22 Pitts, "Gloria E. Anzaldúa's Autohistoria-teoría as an Epistemology of Self-Knowledge/Ignorance," 360.

23 Wittgenstein, *Philosophical Investigations*.

24 Pitts, "Gloria E. Anzaldúa's Autohistoria-teoría as an Epistemology of Self-Knowledge/Ignorance," 365.

25 Brownmiller, *In Our Time: Memoir of a Revolution*, 280-281. Qtd. in Fricker, *Epistemic Injustice: Power and the Ethics of Knowing*, 149-150.

26 Fricker, *Epistemic Injustice: Power and the Ethics of Knowing*, 151.

27 Anzaldúa, "now let us shift … the path of conocimiento … inner works, public acts," 139.

those narratives are not created by them.

I have identified two basic conditions of body narratives: (1) they have to be traversed by first-person experience and (2) by language. But I have not specified the sense in which a body could be a body narrative. Let us start by considering an example of narratives embedded: tattooed bodies.

Kosut (2000) points out that:

> Aside from their obvious visual and aesthetic quality, tattoos also have a distinctive narrative quality. Like every photograph, every tattoo has a story behind it. In what context are tattoo narratives constructed, to whom are tattoos being communicated, and what is being said?[28]

One could think tattoos are narratives like pictures. According to Kosut, however, "In tattoo narratives a sophisticated awareness of the corporeal dimension of existence is articulated."[29] Hence, there is a sense in which tattooed bodies *are* body narratives and not just a narrative on the surface of the body. The body is "culturally constructed and read by others [and by oneself]."[30] Position, visibility, and size (among other factors) are constitutive of the narrative expressed by tattoos.[31] Consequently, a tattooed person usually, though not necessarily,[32] reflects (meaning-making process) how their body was going to express the narrative. Tattoos imply body narratives inasmuch as —and only in cases where— they comprise a meaning-making process in regards to the way in which the body is part of the narrative by expressing it.

There are two interesting corollaries from the tattoo example. Body narratives depend on body modifications and particular meaning-making processes in which the body is constitutive of the expression of the narrative. Heavey (2015) claims that "In analyzing narrative body constructions, [...] as people tell stories that are explicitly or implicitly about their bodies, they are also constructing and performing versions of those bodies."[33] A body narrative then, in Heavey's terms, is the materialization (by a body modification) of a narrative body construction. Heavey's claim raises the question of to what extent body modifications are acceptable. This is indeed a challenging question since several

---

28 Kosut, "Tattoo Narratives: The Intersection of the Body, Self-identity and Society," 82.

29 Ibid., 96.

30 Ibid.

31 Ibid., 85.

32 The tattoo might not reflect a meaning-making process if the person is unconscious during the tattooing.

33 Heavey, "Narrative Bodies, Embodied Narratives," 434.

different cases[34] need to be considered. From the present context, it is modestly suggested to be a simple idea: to judge a body modification, one should consider the narrative body construction that motivates and endorses it.

To illustrate, look at the following two polemic body modifications. First, transgender identity is related to certain body modifications and presentations, which can range from wearing makeup to undergoing cosmetic surgery and/or sex reassignment surgery. Such body modifications can be understood as political acts of resistance, in as much as they manifest diverse meanings of gender in opposition to the oppressive heterosexual and cisgender normativity.[35] Some narratives motivate the same kind of body modifications, but they "foreclose multiple (otherworldly) interpretations of terms like 'woman.'"[36] Such narratives replicate oppressive gender stereotypes: for example, that women have to wear makeup or have large breasts to be considered attractive. In addition, social judgments related to body modifications can foreclose the possibility of a multiplicity of narrative body constructions. For instance, people who are transgender can be coerced to reproduce a pathologizing and binary-exclusive narrative (i.e., the wrong-body model) in order to be allowed to change their gender.[37]

Second, "transabled identity" is related to certain body modifications, such as voluntarily amputating a healthy limb. In a broad sense, "'transability' describes an able-bodied person's need to modify his or her body to acquire a physical impairment/disability."[38] Without taking a stance on the ethical debate about whether it should be permitted or not, the point is that assuming a fixed transabled body narrative could parallel what happened with the transgender wrong-body model. The problem is not that institutionally recognizing transability would entail an explosion of cases —as some have discussed[39]—, but that it would "shape an emerging social identity"[40] by *limiting* the possibilities of narrative construction of the body. Conversely, exploring the narratives behind such desired modification may reveal different interpretations of disability as well as hidden

---

34 Take for instance the following (not comprehensive) list of body modifications/modifiers or actions: amputations, plastic surgery, drug consumption, piercings, physical exercise, eating, and meditation.

35 Bettcher, "Trapped in the Wrong Theory: Rethinking Trans Oppression and Resistance," 387-390.

36 Ibid., 401.

37 Ibid., 402-403; Stone, "The *Empire* Strikes Back: A Posttranssexual Manifesto," 152 ff.

38 Baril, "Needing to Acquire a Physical Impairment/Disability: (Re)Thinking the Connections between Trans and Disability Studies through Transability," 31.

39 Bayne and Levy, "Amputees by Choice: Body Integrity Identity Disorder and the Ethics of Amputation," 85; Johnston and Elliott, "Healthy Limb Amputation: Ethical and Legal Aspects," 434.

40 Ibid., 434.

ableist biases.[41]

The relevant questions should be: What are the narratives that motivate a body modification like voluntary amputation? How are these narratives related to ableist or transphobic biases? Which narratives should be embraced to foster a more inclusive and tolerant society? In short, what is intended to show through these cases is that the focus should not be on the body modification by itself but on the narrative behind it.

CONCLUSION

The first section of this paper aimed to show that the body as a physical object is not epistemologically significant. Considering the physical body as a pertinent factor when assessing cognitive abilities is *typically* unfounded. With the exclusion of some particularly difficult cases, ascribing cognitive inferiority to marginalized bodies is morally wrong. The second section of this paper aimed to show that the experiential body can be a narrative and thus epistemologically significant. As the body determines narrative(in the sense that the body shapes experience), and as it serves as an expression of a narrative by embodying it in physical form, the experiential body indeed becomes a narrative. It serves both as the expression and the source of the narrative.[42] In summary, the body is a fundamental factor in marginalization and can become a narrative.

I will now briefly explore (in a prescriptive spirit) the significance of body narratives in achieving the aim of fostering an inclusive society and establishing moral responsibility. According to Medina (2013), "Both our ability and our inability to relate to others (and to particular aspects of ourselves) is mediated by the social imagination, the kind of imagination that opens our eyes and hearts to certain things and not others, enabling and constraining our social gaze."[43] Given that what people publicly narrate is what society imagines, narratives help to constitute the social imagination. Hence, the role of narratives in the transformation of the social imagination is fundamental. Narratives are sometimes (wrongly) conceived as immaterial and fanciful. As tangible living meanings, body narratives powerfully impact the social imagination. To proudly think of and show one's body (oneself) contributes to the transformation of the social imagination in a good or bad way. This leads to (at least) two important consequences. First, people are responsible for modifying their bodies in order to enhance the social imagination

---

41 Arfini, "Instructions for Becoming Disabled: A Narrative Analysis of the Project of the Transabled Body;" Reynolds, "Toward a Critical Theory of Harm: Ableism, Normativity, and Transability (BIID)."

42 What the experiential body is and what one does with the experiential body are scarcely distinguishable. This mirrors the heading quotation of Merleau-Ponty: "the expression is indistinguishable from the thing expressed."

43 Medina, *The Epistemology of Resistance. Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations*, 22.

and narrate one's identity. Second, marginalized bodies may express resistant narratives to the extent that they endorse an equally resistant narrative body construction. For example, a Black person is expressing a resistant narrative when they proudly show their Afro-textured hair.

Before closing this paper, a clarification is needed. This argument does not imply that everyone has the responsibility to be a marginalized body. Instead, since bodies are narratives that express identities, everyone has the responsibility to ensure such correspondence and to avoid being influenced by oppressive body narratives. Here, I do not intend to advocate for any particular body narrative or modification. The important point is that when considering body modifications, it should be considered: What kind of body narrative is behind the modification?

෴

BIBLIOGRAPHY

Anzaldúa, Gloria. "now let us shift ... the path of conocimiento ... inner works, public acts." In *The Gloria Anzaldúa Reader*, ed. Ana Louise Keating. Duke University Press, 2009.

Arfini, Elisa A. G. "Instructions for Becoming Disabled: A Narrative Analysis of the Project of the Transabled Body (Istruzioni per diventare disabile. Un'analisi narrativa del progetto sul corpo transabile)." *Studi Culturali* 7, no. 3 (2010): 343-364. DOI: 10.1405/33554

Baril, Alexandre. "Needing to Acquire a Physical Impairment/Disability: (Re)Thinking the Connections between Trans and Disability Studies through Transability." *Hypatia* 30, no. 1 (2015): 30-48.

Bakare-Yusuf, Bibi. "The Economy of Violence: Black Bodies and the Unspeakable Terror." In *Feminist Theory and the Body*, eds. Janet Price and Margrit Shildrick. Edinburgh University Press; Routledge, 1999.

Bayne, Tim, and Levy, Neil. "Amputees by Choice: Body Integrity Identity Disorder and the Ethics of Amputation." *Journal of Applied Philosophy* 22, no. 1 (2005): 75-86.

Bettcher, Talia Mae. "Trapped in the Wrong Theory: Rethinking Trans Oppression and Resistance." *Signs* 39, no. 2 (2014): 383-406.

Brownmiller, Susan. *In Our Time: Memoir of a Revolution*. Dial Press, 1990.

Fricker, Miranda. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press, 2007.

Garland-Thomson, Rosemarie. "The Case for Conserving Disability." *Bioethical Inquiry* 9 (2012): 339-355. DOI: 10.1007/s11673-012-9380-0

Heavey, Emily. "Narrative Bodies, Embodied Narratives." In *The Handbook of Narrative Analysis*, eds. Anna De Fina and Alexandra Georgakopoulou. 2015. DOI: 10.1002/9781118458204.ch22

Husserl, Edmund. *Experience and Judgment.* Trans. James Churchill and Karl Ameriks. Routledge & Kegan Paul, 1973.

Johnston, Josephine, and Elliot, Carl. "Healthy Limb Amputation: Ethical and Legal Aspects." *Clinical Medicine* 2, no. 5 (2002): 431-435.

Kosut, Mary. "Tattoo Narratives: The Intersection of the Body, Self-identity and Society." *Visual Sociology* 15, no. 1 (2000): 79-100. DOI: 10.1080/14725860008583817

Medina, José. *The Epistemology of Resistance. Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations.* Oxford University Press, 2013.

Merleau-Ponty, Maurice. *Phenomenology of Perception.* Trans. Colin Smith. Routledge & Kegan Paul, 1962.

Okruhlik, Kathleen. "Gender and the Biological Sciences." *Canadian Journal of Philosophy* 24, Sup. 1 (2013): 21-42. DOI: 10.1080/00455091.1994.10717393

Pitts, Andrea. "Gloria E. Anzaldúa's Autohistoria-teoría as an Epistemology of Self-Knowledge/Ignorance." *Hypatia* 31, no. 2 (2016): 352-369.

Reynolds, Joel Michael. "Toward a Critical Theory of Harm: Ableism, Normativity, and Transability (BIID)." *APA Newsletter on Philosophy and Medicine* 16, no. 1 (2016).

Scully, Jackie Leach. "From 'She Would Say That, Wouldn't She?' to 'Does She Take Sugar?' Epistemic Injustice and Disability." *International Journal of Feminist Approaches to Bioethics* 11 (2018): 106-124. DOI: 10.3138/ijfab.11.1.106

Stone, Sandy. "The *Empire* Strikes Back: A Posttranssexual Manifesto." *Camera Obscura: Feminism, Culture, and Media Studies* 10, no. 2 (1992): 151-176.

Wittgenstein, Ludwig. *Philosophical Investigations.* Macmillan Publishing Company, 1953.

ACKNOWLEGEMENTS

RESOLVING THE PARADOX OF PLEASURE IN ARISTOTLE'S *NICOMACHEAN ETHICS* VII & X

by Rashad Rehman

The University of Toronto

ABSTRACT ‖ *Many read Aristotle as having two inconsistent accounts of pleasure – what G. E. L. Owen called the 'A' and 'B' accounts of pleasure. The A account holds that pleasures are "activities" (*energeiai*) (NE, VII; EE, IV, 1153a9-15) and the B account holds that pleasures complete or perfect* energeiai, *but are not themselves* energeiai *(NE, X, 1174b14-75b1). Specifically for Owen, a reconciliation of the A-B dilemma involved treating A and B, respectively, as two, mutually exclusive* accounts *of pleasure. A and B, qua* accounts, *perform the same task: to specify the nature of pleasure. However, while I reject the impetus to offer a solution to the A-B dilemma as Owen construed it – in following with the recent literature – I nonetheless preserve Owen's formulation of 'A and B.' I argue that the 'A' and 'B' constitutive of the A-B dilemma are individually revisable such that, under a weakening of the claims in both A and B – that is, A and B are more plausibly* explanations *and not* accounts *(in Owen's sense) of pleasure – we can achieve a plausible, Aristotelian account of both A and B which makes them consistent.*

"What a strange thing that which human beings call 'pleasure' seems to be…" —Plato, Phaedo, 60b4-5, translated by G.M.A. Grube (with modification)

I. INTRODUCTION

G.E.L. Owen (hereafter 'Owen') (1922-82), among others, read Aristotle as having two inconsistent *accounts* of pleasure – what he called the A and B accounts of pleasure.[1] The A account holds that pleasures are "activities" (*energeiai*) (*NE*, VII; *EE*, IV, 1153a9-15) and the B account holds that pleasures complete or perfect *energeiai*, but are not themselves *energeiai* (*NE*, X, 1174b14-75b1). For Owen, a reconciliation of the A-B dilemma involved treating A and B respectively as two, mutually exclusive accounts of pleasure. A and B, *qua* accounts, perform the same task: to specify the nature of pleasure. While I reject the impetus to offer a solution to the A-B dilemma as Owen construed it – in following with the recent literature – I nonetheless preserve Owen's formulation of 'A and B.'[2] I argue that the 'A' and 'B' constitutive of the A-B dilemma are individually revisable such that, under a weakening of the claims in both A and B – that is, A and B are more plausibly *explanations* and not *accounts* (in Owen's sense) of pleasure – we can achieve a plausible, Aristotelian account of both A and B which makes them consistent. The paper relies on a subtle distinction (that I articulate) between the *interpretive* dilemma and *philosophical* dilemma of pleasure in Aristotle. While Owen addressed the interpretive question, I will be invested (i) in a counterfactual solution to the interpretive dilemma, namely, a solution which shows how A and B could have been consistent if Aristotle had weakened (i.e., in the sense of making a lesser claim) A and B

---

1 Four brief, prefatory notes. First, all references to Aristotle – unless specified otherwise – will be from The Complete Works of Aristotle. 2 vols. Ed. Jonathan Barnes. (Princeton: Princeton University Press, 1984). Please also note that all Greek words are transliterated into English, and all abbreviations are summarized at the end of this paper. Second, throughout the paper I will use Owen's translations of Aristotle's Greek, although when appropriate I suggest alternative translations. The standard translations are as follows: activity (*energeia*), pleasure (*hedone*), pleasures (*hedonai*), state (*hexis*), proper (*oikeia*), complete, perfect (*teleein*), origin, source (*genesis*), movement, motion (*kinesis*). Third, following a citation of Aristotle using the (abbreviated) title of the book, chapter and section number, I will then truncate the citation with only Becker numbers (referring to the previously cited work). Fourth, I will not be directly invested in Aristotle's approach to the heterogeneity problem of pleasure in this paper; that is, the question of what feature (if any) of objects and/or our phenomenology is common to all, or shared among, experiences of pleasure. One possible implication of my paper, however, might be a rejection of the problem altogether since I abandon the prospects of, or suggest an alternative approach to, a reconciliation of Aristotle's two accounts of pleasure. Whether one accepts my conclusion or not (and the implications for the heterogeneity problem), one should still consult Matthew Strohl's excellent "Aristotle on the Heterogeneity of Pleasure" in Pleasure: A History. Ed. Lisa Shapiro. (Oxford: Oxford University Press, 2018), 42-65.

2 Verity Harte's "The Nichomachean Ethics on Pleasure" in The Cambridge Companion to Aristotle's Nicomachean Ethics. Ed. Ronald Polansky (New York: Cambridge University Press, 2014), 312.

---

as I suggest and (ii) the philosophical merits of my reconfigured, Aristotelian solution to the A-B dilemma.

With the foregoing description in mind, the argument is as follows. First, although Aristotle's A-B are straightforward definitions of pleasure (aiming to account for the heterogeneity of pleasure), A and B themselves are more plausible when weakened as explanations (not accounts) one might provide of pleasure. A distinction between accounts and explanations is needed here: Owen's understanding of A and B are two accounts or logoi of pleasure; that is, they each offer to explain the nature of pleasure and therefore cannot both be true. Explanations (*aitiai*) for Aristotle can complement one another, however, and have the power to elucidate different features of the same phenomena.[3] Second, recall that for Aristotle each being has four possible explanations (*aitiai*) (*Phys.* II.3, 194b16-195a1): material explanation ('that out of which a thing comes to be'), final explanation ('end or that for the sake of which a thing is done'), formal explanation ('the form' or 'definition of the essence'), and efficient explanation ('the source of the primary principle of change or stability'[4]). On my weakened reconfiguration of the A-B distinction, A refers to the material explanation of pleasure i.e., as an activity (pleasure is an *energeiai*). B refers to the final explanation i.e., to perfect *energeiai*. (I also classify 'being unhindered' as the formal explanation, that is, 'the form' or 'definition of the essence', and efficient explanation 'that out of which a thing comes to be', or the art of performing the activity in question – I use the example of 'decorating' in the paper). In the end, I conclude that if my deflationary, Aristotelian account of A-B is correct, there is another successful candidate among Owen's options for making A-B consistent. While such a reconfiguration of the A-B dilemma causes my revision of the dilemma to be less of a solution and more of an Aristotelian reconfiguration of the dilemma, I will explain why this is a negligible feature of my account.

II. OWEN, SETTING UP THE PROBLEM AND THE A-B DILEMMA

In his 1971 "Aristotelian Pleasures"[5], Owen wrote the following: "I have in mind the more baffling inconsistency that seems to lie at the heart of the discussions on pleasure on which philos-

---

3 Throughout this paper, I will use 'account' instead of 'definition'. By 'account', I mean – and take Owen to mean – the original (Platonic) sense of logos, that is, an account of some phenomenon x such that (analytically) if you have two non-equivalent (or identical) logoi to explain x, then the two logoi explaining x are necessarily incompatible. Depending on the kind of definition one offers, however, definitions can be inconsistent in defining the same explanandum in a way accounts (as I have defined it) cannot be, e.g., analytic definitions are consistent with ostensive definitions. To avoid these problems, I continue with 'account.'

4 Terence and Fine translation.

5 Owen, G.E.L. "Aristotelian Pleasures" Proceedings of the Aristotelian Society, 72 (1971-72): 135-152

ophers have chiefly drawn", namely, "the studies that now appear in the seventh and tenth books of the Nicomachean Ethics."[6] According to the traditional debate, Aristotle apparently has two answers or accounts (A and B, respectively) (*NE*, VII.11, 1152b8-12).[7] However, before making the formal distinction between A and B, Owen makes two prefatory remarks. First, from the (apparent) inconsistency of A and B, his goal is a unified account of A and B: "...[Aristotle] is looking for some common feature in all pleasures, some necessary and sufficient condition for anything to be enjoyable; and he uses something like a method of concomitant variations to isolate it. And his critics have long complained that this search for the unit is a delusion."[8] However, a distinction between two kinds of the A-B dilemma will facilitate understanding Owen and what is constitutive of a solution:

Interpretive Dilemma: The question of whether one conceptualization of Aristotle's account(s)/

---

6   Ibid., 135. Two comments. First, Owen, like myself, is not primarily invested in the historical question of Aristotle's position, namely, how it was handled in antiquity. While Owen makes lighthearted remarks in that direction, it is worthwhile to note (many) points of contact with Plato, e.g., ibid., Harte, 288-318. Gerd van Riel's "Aristotle's Definition of Pleasure: A Refutation of the Platonic Account" Ancient Philosophy 20.1 (2007): 119-138. Second, post-Owen, the mistake of conflating the A and B persists, e.g. in Dana LaCourse Munteanu's "Aristotle on Pleasure and Learning": "Aristotle generally defines pleasure as an activity and end (Nicomachean Ethics 7.1153a10 = Eudemian Ethics 6). But pleasures complete activities without, in themselves, being activities (Nicomachean Ethics 10.1174b–1175a). Thus, pleasure is described as a completion of an activity: 'as a supervening end' (Nicomachean Ethics 10.1174b32)." Encyclopedia of the Sciences of Learning. Ed. Norbert M. Seel. Boston: Springer, 2012). Even if one is convinced that A and B express the same content (or are reconcilable such that they end up expressing the same content), one still should recognize the perplexity of the A-B dilemma.

7   Some scholars post-Owen reject Owen's set-up at the outset. For example, Aristide Tessitore in "A Political Reading of Aristotle's Treatment of Pleasure in the Nicomachean Ethics" argues that "Aristotle's treatment of pleasure in Book X abstains from questions about the nature of pleasure itself. Whereas Book VII attempts a definition of pleasure (even an incomplete or tentative one), Book X ignores this question altogether and appears content to confine itself to a description of pleasure. Despite its more systematic character, Book X brings us no closer to answering the question: What is pleasure?" Political Theory 17.2 (1989): 253. The problem with Tessitore's concession here is twofold. First, the definition/description distinction is not textually plausible. This was highlighted in pre-Owen scholarship, e.g., G.B. Kerferd's review (The Classical Review: 10.2 (1960): 118-120) of Godo Lieberg's Die Lehre von der Lust in den Ethiken des Aristoteles. (Munich: Beck, 1958) as well as post-Owen scholarship, e.g., ibid., Harte, 309. Aristotle, to all appearances, uses the language of identity (eimi) in his discussion of pleasure in both *NE*, VII and X. Second, Owen defended his position that A and B were accounts of pleasure and hence Tessitore's distinction requires more motivation (see my articulation of A-B below).

8   Ibid., 142.

explanation(s) of pleasure is (i) found in Aristotle's writings (ii) is accurate and (iii) textually/exegetically defensible.

Philosophical Dilemma: The question of whether the account(s)/explanation(s) of pleasure found in Aristotle's writings are defensible philosophically in their own right.[9]

Second, the processes which restore us to a natural state are only incidentally pleasant (*NE*, VII.12, 1152b34-35).[10] 'X is incidentally y' is defined by Aristotle as follows: x's are not always or necessarily, or even usually, y's. As I understand Owen, an x that is y is only incidentally y when it is not the nature of x's to be y's or it is not because of x that this one is y (Top, I.5,1026b37-27a8, 1025a28-29). I will now explain the A-B account as Owen articulates it.

The A account is that pleasures are *energeiai* (*NE*, VII; *EE*, IV, 1153a9-15) (this account is offered in *EE*, VII, 11-14).[11] Here is the thesis (what Owen designates as 'T'): Pleasures are the unhindered activities of our natural faculties/states [hexis] (*NE*, VII.12, 1153a14-15). The result of the unhindered activity of our natural faculties/states is pleasure. The best life is some pleasure or class of pleasures (1153b7-13). We do not enjoy a process e.g., relieving hunger.[12] A is not concerned – minus 1153a20-23 – with the relation between the enjoying and what is enjoyed[13]; rather,

---

9   With respect to the interpretive dilemma it is important to distinguish exegetically defensible from textually defensible. As I will explain below, my counterfactual solution to the interpretive dilemma is not exegetically defensible but textually defensible; that is, the justification of my use of Aristotle is rooted in the text of Aristotle, even if not an exegesis of him i.e., in articulating the view he himself adopted and defended.

10   Ibid., 144.

11   Two notes. Second, the translation of *energeia* is Owen's, and is not without its difficulties. Francisco J. Gonzalez in his "Aristotle on Pleasure and Perfect" reports the difficulty: "Energeia is here difficult to translate. It could be rendered 'actuality', but actuality is too static a notion to do justice to the dynamic character of *energeia*. The translation 'activity', on the other hand, makes *energeia* too much like a process or motion", conceding that "for want of a better translation" "activity" is the most appropriate translation of *energeia*. Phronesis 36.2 (1991): 145ff8. On a more modest reading of Gonzalez, the problem with "activity" is only that it is easily confused with "process" (which Aristotle explicitly wants to avoid (*NE*, VII.12), not that "activity" itself is a bad translation per se. The (Ancient) Greek suffix '-ia' is ambiguous in this way: it can either mean (1) process/condition or (2) state/condition (although Aristotle distinguishes between these in his discussion of 'quality' cf. Cat. 8). Gonzalez does not consider 'actualization' (thank you to Mohan Matthen for this suggestion) or 'activation.' Activation is not an atypical translation for *energeiai* e.g., Lloyd Gerson's translation of Plotinus' Enneads. (Cambridge: Cambridge University Press, 2019), 905. (Following Owen, I have used transliterations of Aristotle's Greek for larger readability among a broader audience).

12   Ibid., 138.

13   Ibid., 143.

A is about what enjoying is.[14] For example, "gaming is one of my pleasures" is an A-usage; that is, the English gerundive 'gaming', just is one of my pleasures.[15] The question "what is pleasure?" therefore means "what is the character of a pleasure, what ingredients of a life are enjoyable in themselves?"[16] The nature of enjoying does not come into question in A: "If we are asked what is admirable about politics or detestable about beagling, we do not stop to ponder what admiring or detesting is before embracing the question." [17]

The B account is that pleasures complete or perfect *energeiai*, but are not themselves *energeiai* (*NE*, X, 1174b14-75b1). It is an open question whether we choose the best life for the sake of pleasure or pleasure for the sake of the best life.[18] Aristotle rejects here that the Greek equivalents of these verbs (taking pleasure, enjoying) have the logic of process-verbs.[19] Consistent with this is J.O. Ursom's thesis that for Aristotle, "different activities are differently enjoyable"[20], every act has its own proper (*oikeia*) pleasure. B is not concerned with the misidentification of the activity enjoyed[21], it simply states that pleasure augments the activity enjoyed.[22] Pleasure is an end-in-itself, not identifiable with the finest object or the well-conditioned faculty.[23] Pleasure is not the enjoyed activity.[24] For example, "I get pleasure from (or take pleasure in) gaming," is an example of a B-usage. "Gaming" is not itself the pleasure (*qua energeia*), but pleasure is found in the perfection of "gaming."[25]

---

14 Ibid., 147.

15 Ibid., 142. Although Owen writes that on A "what is pleasure?" amounts to "what is the character of a pleasure, what ingredients of a life are enjoyable in themselves?" (142), the point is that on A, there persists the question of the heterogeneity of pleasure: the question of what is common to all pleasures. While Owen's inference from A is that the nature of enjoyment (or 'taking pleasure in') is not the central question, his definition of A clearly marks the heterogeneity question.

16 Ibid., 142.

17 Ibid., 142.

18 Ibid., 136.

19 Ibid., 138.

20 Ibid., 143.

21 Ibid., 144.

22 Ibid., 145.

23 Ibid., 146.

24 Ibid., 146.

25 A question that Owen does not consider is the question of whether *energeiai* can perfect *energeiai*, such that *energeiai* are both activities and able to perfect other *energeiai*. While Owen does not provide an answer to this question explicitly, he has theoretical motivation to avoid it: if it were true that *energeiai* could perfect other *energeiai*, Aristotle would be hard-pressed to choose either *energeiai* or a perfected-*energeiai* as his definition of pleasure (for they cannot both be a definition of pleasure). A similar worry is found in Richard Kraut: "...perhaps Aristotle is merely trying to avoid a possible misunderstanding: when he says that pleasure completes an activity, he does not mean that the activity it accompanies is in some way defective, and that

III. WHAT IS THE DIFFERENCE BETW*EE*N A AND B? SOME ATTEMPTS AT RECONCILIATION

Since they agree with the core presupposition of the traditional A-B dilemma, namely, that A and B are mutually exclusive, inconsistent accounts of pleasure, Owen considers various attempts to reconcile them.[26] While some of the attempted reconciliations are better than others, Owen concludes that none succeed in presenting a successful solution to the A-B dilemma. For example, "perhaps it occurred to Aristotle that the activities of the natural states which served as A's paradigms of pleasure need not be enjoyable at all."[27] He gives the example of Smith who is exercising his wits on an argument; however, "his wits are blunt, he is tired, the argument is tangled... so B is spelling out the further conditions that are requisite for pleasure—sharp wits, impeccable object."[28] Owen explains that this will not work in explaining the difference: First, "...such conditions might be covered by A's requirement that the activity proceed unhindered."[29] Second, "B does not conclude that when such conditions are satisfied the activity is a pleasure, only that pleasure inevitably ensues."[30] Third, "B seems curiously unaware of the central claim of A that only self-contained activities and not end-directed

---

the pleasure improves the activity by removing this defect. Aristotle's language is open to that misinterpretation because the verb that is translated 'complete' (teleein) can also mean 'perfect'. The latter might be taken to mean that the activity accompanied by pleasure has not yet reached a sufficiently high level of excellence, and that the role of pleasure is to bring it to the point of perfection...Taking pleasure in an activity does help us improve at it, but enjoyment does not cease when perfection is achieved—on the contrary, that is when pleasure is at its peak. That is when it reveals most fully what it is: an added bonus that crowns our achievement." "Aristotle's Ethics", The Stanford Encyclopedia of Philosophy (Summer 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2018/entries/aristotle-ethics/>. As I interpret Kraut, we might say that Aristotle is right that pleasure is an activity (A), but also that its perfection results in a higher degree of pleasure (B). While this would solve the A-B dilemma's core worry with respect to consistency, it comes with a cost: B does not merely associate pleasure (let alone a higher degree of pleasure) with the perfection of an activity (as though it were a contingent relation), but identifies pleasure with the perfection of an activity.

26 Reconciliation of Aristotle on pleasure is also significant for other, related aspects of Aristotelian scholarship e.g., Devin Henry "Aristotle on Pleasure and the Worst Form of Akrasia" Ethical Theory and Moral Practice 5.3 (2002): 255-270 (action theory); Amelie Oksenberg Rorty "The Place of Pleasure in Aristotle's Ethics" Mind 83.332 (1974): 481-497 (ethics); David Bloch's "Aristotle on The Pleasure of Learning" Classica et Mediaevalia 64 (2013): 161-174 (learning).

27 Owen, "Aristotelian Pleasures", 146.

28 Ibid., 146.

29 Ibid., 146.

30 Ibid., 146.

processes are enjoyable..."[31] While Owen eventually concedes that we ought to "leave the attempt to build a bridge from A to B and consider a declaration of independence"[32], he first evaluated various attempts at putting A and B together, explaining why none of the options achieves a plausible A-B reconciliation. For brevity's sake, I here summarize the attempted reconciliation with Owen's reply (Figure 1):

Reconciler	Attempted Reconciliation	Owen's Objection
Festugiere, Gauthier-Jolif, Dirlmeier, Stewart	A is hastier and more polemical than B; A is simply a finished version of B.	This claim is exaggerated: both A and B offer explicit, positive accounts of pleasure.
Hardie	Conflates A and B in suggesting that pleasure is an activity or the completion of an activity.	If you conflate A and B, there is no explanation for why in A and B we find different accounts of pleasure.
Ross	If we dispute A and B (in his language "contradiction [between accounts]"), preference must be given to *NE*, X.	There is also a positive account in B – so priority cannot be given to A because B offers a positive account of pleasure.
Stewart	A and B are essentially the same.	This overlooks the root trouble of A and B, namely, that two different accounts are offered.
Owen	A focuses on genesis (1152b13, 23) and B on kinesis (1174a19-b9). A and B do not have the same target.	This does not identify the root of the trouble between A and B.

Figure 1. Attempted Reconciliations with Owen's Replies.

Regarding these attempts at reconciliation as unsuccessful[33], Owen performs a linguistically dense analysis of Aristotle's Greek verbs and nouns about pleasure before concluding his paper. He gives the following list: hedone: Aristotle's main verb for pleasure, apolausis: alternative noun for hedone, terpsis: alternative noun for hedone, chairein: associate verb (more common than hedesthai in *NE*), hedesthai: associate verb, areskein: associate verb, terpein: associate verb, and agapan: associate verb.[34] Owen infers that "at various cardinal points in B, but never in A, the argument turns directly on an appeal to

the behaviour of the verb"[35], and then he finishes the paper with how this plays out.[36] I leave his findings to a footnote for one's own assessment, since I am concerned not merely with Owen's position per se, but with the core presupposition of A-B.[37]

Owen concludes:

> When Aristotle rejects the thesis that pleasure is a process in A, he is offering to tell us what our real pleasures are, what is really enjoyed or enjoyable. When he rejects a thesis in the same form of words in B, he is offering to tell us what the nature of enjoying is by reviewing the logical characteristics of pleasure-verbs. In B he moves naturally to question what enjoying contributes to the enjoyed activity, and apart from one peripheral hint there is no sign of that question in A...the question becomes then: what can be said about enjoying X-ing that cannot be said about X-ing, and the converse? Verbs and their adverbs, and then their other logical features, take the center of the inquiry; and for the philosophically suspect enterprise of A are substituted the admirable studies of B.[38]

The problem remains, however, as Aristotle is clearly offering accounts of pleasure. Since this problem persists, my entry into the discussion will take the form of a revaluation of the presupposition of the defenders of the traditional A-B debate; namely, that the A-B dilemma should be regarded as accounts of pleasure, each individually attempting to explain the nature of pleasure by means of a definition.

IV. THE A AND B AS LOGOI OR AITIAI?

While Owen addressed the interpretive dilemma of A-B, I propose a counterfactual solution to the interpretive dilemma, namely, a solution which shows how A and B could have been consistent if Aristotle had weakened A and B. Owen's formulation of the A-B dilemma requires that both A and B are 'accounts' of pleasure. By 'account', I simply

---

31 Ibid., 147.

32 Ibid., 147.

33 While I cannot defend at length Owen's rejection of these reconciliations, it is worth noting that soon after Owen's publication (in the same decade) other objections against them were also raised e.g., Philip Webb's "The Relative Dating of the Accounts of Pleasure in Aristotle's Ethics" Phronesis 22.3 (1977): 235-262; Frank Lucash's "More Pleasure in Aristotle" Rivista di Filosofia Neo-Scholastica 66.1 (1974): 126-130.

34 Ibid., 147.

---

35 Ibid., 147.

36 Ibid., 148-151.

37 In 1173a15-22, Aristotle recognizes and finds parallels for one feature of the logic of pleasure-verbs, namely, that they are degreed, i.e., you can be more or less pleased. In 1173a31-b4, Aristotle says that we get to be pleased quickly or slowly, but not be pleased quickly or slowly. Process verbs collect adverbs of relative speed. In 1173b7-13, perhaps here B is A's interpretation that pleasure is a kinesis. A asks what is enjoyed or enjoyable, and is accordingly ready to argue: a bodily function. But B asks what enjoying is, and, by considering the logical requirements on subjects for enjoyment-verbs, replies: not a bodily function. In 1174b7-9, being pleased can happen in an instant. In 1174a13-b7, Aristotle contrasts pleasure with seeing: "There is no reason to deny that Aristotle is interested in the possibility of interrupting some process without falsifying the natural description of what is interrupted. And "enjoying" can never stand for such a process." (Ibid., 150) These conjunctively (or cumulatively) show that Aristotle is invested in the logic of enjoyment-verbs (Ibid., 151). For criticism of Owen's analysis of the Greek verbs, see ibid., Harte, 310ff35.

38 Ibid., 151.

mean the original (Platonic) sense of logos, that is, an account of some phenomenon x such that (analytically) if you have two non-equivalent (or identical) logoi to explain x, then the two logoi explaining x are necessarily incompatible.[39] I am not making the strong claim that Owen's understanding of A and B as accounts is Aristotle's rendition of logoi; instead, I merely point out that according to the traditional set-up, pleasure is either A or it is B (or perhaps neither A nor B), but it is not both A and B, unless a specification for A and B's reconciliation can be successfully offered. Owen does this by setting up A and B to be inconsistent with one another (recall "the residual problem remains intractable"), that is, they are not complementary logoi and specifying A and B as essentially competing definitions of pleasure. Here I would like to contrast this sense of logos with Aristotle's notion of 'explanation' (aitia) found in *Phys.* II.3. While logoi are inconsistent with one another and hence cannot both be true simultaneously, *aitiai* are complementary since they explain different aspects or features of the same phenomena. For Aristotle, there are four *aitiai* (194b24-195a1):

(1) Material Explanation: 'that out of which a thing comes to be and which persists'
(2) Formal Explanation: 'the form or the archetype, i.e., the definition of the essence'
(3) Efficient Explanation: 'the primary source of the change or rest'
(4) Final Explanation: 'end or that for the sake of which a thing is done'

Consider a house. What explanation can there be of a house? I can explain the home in terms of its constitutive materials (material explanation), the builders of the house (the efficient explanation), the idea or blueprint of the house (formal explanation) or the purpose for the house being built (final explanation). Neither of these explanations are individually exhaustive, and they do not conflict with one another.[40] However, since A

---

39 While I use logos in a specific sense, there are many other uses in Aristotle e.g., Omer Aygun's The Middle Included: Logos in Aristotle. (Evanston: Northwestern University Press, 2017). I do concede that my exegesis of Aristotle's notion of aitia might be much more rigorous than that of logos in Aristotle, such that the latter deficit makes my view more implausible or at any rate incomplete. In reply, while the objection is right to highlight my emphasis on the former, I do not see the problem of incompleteness in the case of the latter. My view does not hinge on a studied view of logos in Aristotle i.e., an account of logos found in Aristotle's writings; however, my argument does hang on a correct appraisal of Aristotle's *aitiai*. My argument puts Aristotle's four *aitiai* to philosophical work inasmuch as my argument is Aristotelian-motivated; however, my definition of logos need not be Aristotelian: all my argument requires is that it is an analytic truth that if you have two non-identical logoi for one and the same phenomenon, then the two logoi are necessarily incompatible (and that is a problem for the traditional A-B dilemma).

40 It is appropriate in this context to explain why I use the translation explanation over cause. In ordinary English, inquiring into "the cause of x" usually, if not always,

---

and B are traditionally set up as inconsistent with one another, they are generally taken to be accounts (logoi), not explanations (*aitiai*). I argue that a plausible, Aristotelian account of pleasure is achievable by construing A and B as *aitiai* and not logoi. This requires clarification. First, my argument is not that Aristotle himself thought that A and B were explanations of pleasure, respectively; rather, my argument is that his position is more plausible when A and B are construed as *aitiai*. Second, my argument preserves the Aristotelian impetus of A and B while rejecting the strong, traditional set-up of the A-B dilemma as two competing logoi.

V. THE A AND B AS *AITIAI* DEFENDED

I will now proceed to defend the philosophical merits of the counterfactual solution to the interpretive dilemma of A-B. My (weakened and Aristotelian) reconfiguration of the A-B distinction claims that A refers to the material explanation of pleasure, 'that out of which a thing comes to be', or an activity (*energeiai*), and that B refers to the final explanation, the 'end or that for the sake of which a thing is done', namely, to perfect *energeiai*.[41] I give three motivations for this position. First, following Owen, I think A and B are not consistent or plausible logoi for two reasons: (1) A and B offer two competing definitions of pleasure which – if they are definitions – must either explain one of the horns of the dilemma (as Ross does, cf. Figure 1) or reject one of those horns (cf. Stewart, Figure 1), or A and B must be re-interpreted altogether (as Owen, who thinks B is a much more convincing enterprise).[42] I do not think rejecting or explaining away either A or B is plausible (since Aristotle clearly held both A and B), nor do I think further revision of Aristotle's A and B helps to make them consistent (Owen tried this solution – cf. Figure 1 – but this did not work). By means of a disjunctive syllogism, I conclude that re-interpretation of A and B is a viable option. (2) Since re-interpretation is a viable option against alternatives (cf. Figure 1), it is at least possible to mend the A-B dilemma as merely two explanations of the same phenomenon, namely, pleasure.

---

designates an inquiry which asks what brought x about, that is, the sufficient condition for x's occurrence e.g., "who caused this ruckus?", "what caused the accident?", "what causes something to grow?", et cetera. Take the ruckus example. As I read Aristotle, to give the "why?" of a "ruckus", one must explain not merely who was responsible (efficient cause, the desired answer of "who caused this ruckus?"), but also the material, formal and final cause of the ruckus. Hence, using "cause" in Aristotle's context is likely misleading to English-users e.g., non-efficient causes appear to not be "causes" at all.

41 I also classify 'being unhindered' as the formal explanation, that is, 'the form' or 'definition of the essence', and efficient explanation as 'that out of which a thing comes to be', that is, the art of performing the given activity – to clarify, I use the example of 'decorating' later.

42 Ibid., 151.

Second, using the four *aitiai* of *Phys.* II.3 preserves the Aristotelian impetus of A and B, and hence does not abandon the original Aristotelian impetus of unifying A and B: they are both consistent and unified in being two kinds of explanations of pleasure. Third, reading A and B as *aitiai* makes Aristotle's account (more) plausible. I will now articulate and defend this position.

For example, take "the pleasure of setting up Christmas decorations". I will now explain how my suggestion for reconfiguring the A-B dilemma in terms of the four *aitiai* of Aristotle works. Here is how I map them out using the four *aitiai* (194b24-195a1) and my example:

(1) Material Explanation: 'that out of which a thing comes to be and which persists'
    Decorating (activity) (A)
(2) Formal Explanation: 'the form or the archetype i.e., the definition of the essence'
    Being unhindered [in doing an activity]
(3) Efficient Explanation: 'the primary source of the change or rest'
    The art of decorating
(4) Final Explanation: 'end or that for the age of which a thing is done'
    Perfect or complete activities (B)

In this example, "setting up Christmas decorations" is the activity. In giving an explanation for why I received pleasure from setting up Christmas decorations, we can refer to the activity – the "decorating" – as the material explanation (A). "Decorating", the activity, is the matter or material with which my pleasure is concerned. This is plausible for two reasons. First, it does not reduce pleasure merely to the activity of decorating, as the traditional formulation of the A-B dilemma requires i.e., A requires that a pleasure is an activity.[43] Second, in picking out the material explanation of my pleasure,

---

43 One worry with making 'decorating' the material explanation of pleasure is that it makes the latter appear to be a separable entity. For the sake of brevity, my claim about the material explanation of pleasure – and my *aitiai* interpretation more generally – appears to imply the Separability Thesis: pleasure is separable from, and therefore not immanent in, an activity. Recall the example that Aristotle gives regarding material explanations: "...the bronze of a statue, the silver of the bowl, and the genera of which the bronze and the silver are species." (*Phys.* II, 3). On my view, pleasure is immanent in an activity (e.g., decorating) in the same way that (using Aristotle's example) silver is immanent in a bowl. Notice that in Aristotle's notion of a material explanation in *Phys.* II,3, he makes no explicit claim whether the silver is separable: he only asks what the matter/material explanation of the bowl is. As I read Aristotle, the question of separability follows from specificity. For example, considering only "the bowl", the bowl's being silver is an accidental (non-essential) property of the bowl (Cat. 8) and hence is separable e.g., silver as a property of the

it seems that there is no competing "material" of the activity that is more closely related to my pleasure i.e., "decorating" appears to be the most proximate, plausible activity to pick out, e.g., nothing else serves the same material role in explaining the pleasure. With respect to the formal explanation, the "decorating" is an instance of pleasure inasmuch as the exercise of my faculties to decorate (the activity) was unimpeded (other priorities not competing for my attention, such as cats not destroying the decorations, my body functioning properly to achieve the desired ends, et cetera) (*NE*, VII.12, 1153a15). The claim here is that the explanation of the specific form of the activity is being unimpeded. The efficient explanation I have identified as "the art of decorating." While this might appear to pick out the wrong efficient explanation, e.g., ordinarily Aristotle refers to agents who bring about effects (*Phys.* II.3, 194a31), it is consistent with Aristotle's account that efficient explanations refer to the art of the activity of the agent rather than the agent themselves e.g., 195a6-8; Meta. 1013b6-9.[44] Finally, in explaining the end

---

bowl is non-essential. In offering a material explanation of specifically "the silver bowl", however, the silver is an essential property of the bowl e.g., without the property of being "silver", there fails to exist "a silver bowl." Similarly, when I specify decorating (the activity) as the material explanation of pleasure, I am specifying akin to the latter bowl example: pleasure is not separable from the activity of decorating (the material explanation of pleasure). Pleasures are both activities and ends (*NE*, VII, 1153a10) and arise inseparably when we exercise a faculty (1153a11). Further, in *NE*, X, there is also evidence that although pleasure "supervenes as the bloom of youth does on those in the flower of their age" (1174b32-33), Aristotle rejects the Separability Thesis. The evidence for this is masked by a difficulty in translation. In the Oxford text, we read: "...the pleasure will be involved in the activity..." (1175a1). However, the term "involved" does not appear in Aristotle's Greek. The sentence literally reads: "...there will be pleasure in the activity." (estai en te *energeia* he hedone). The proclitic (preposition) te in Aristotle's Greek means "in", not "involved in." While it does not logically follow that because x is in y that therefore x is not separable from y, Aristotle adds that "without activity pleasure does not arise" (1174b20-21). I conclude that from (i) an activity being a necessary condition for the existence of pleasure and (ii) pleasure residing in the activity itself, it follows that pleasure is not separable from an activity, even if pleasure completes an activity (1174b23).

44 As Andrea Falcon explains: "...an adequate explanation of the production of a statue requires also a reference to the efficient cause or the principle that produces the statue. For Aristotle, this principle is the art of bronze-casting the statue (*Phys.* 195 a 6–8. Cf. Meta. 1013 b 6–9). This result is mildly surprising and requires a few words of elaboration. There is no doubt that the art of bronze-casting resides in an individual artisan who is responsible for the production of the statue. According to Aristotle, however, all the artisan does in the production of the statue is the manifestation of specific knowledge. This knowledge, not the artisan who has mastered it, is the salient explanatory factor that one should pick as the most accurate specification of the efficient cause (*Phys.* 195 b 21–25). By picking the art, not the artisan, Aristotle is not just trying to provide an explanation of the production of the statue that is not dependent upon the desires, beliefs and intentions of the individual artisan; he is trying to, instead, offer an entirely different type of explanation–

(telos) of decorating, the final explanation explains my activity in terms of bringing about pleasure.[45] A worry is that if pleasure is the activity of decorating, what does the end (telos) of decorating – if it is also identified as pleasure – add or contribute to the activity itself? This false dichotomy is rejected by Aristotle "...pleasures are...activities and ends" (*NE*, VII.12, 1153a10), and is alleviated in two steps. First, if I am justified in re-casting A and B as *aitiai*, it is not necessarily true that pleasure is an activity per se, nor is it true that pleasure is merely the perfection/completion of an activity. Second, if A and B are aitai, they do not compete with one another: A explains the material of pleasure, and B explains the finality of pleasure.[46]

CONCLUSION

In this paper, using the set up of Aristotle's two accounts of pleasure in his *NE*, VII and X, I argued that although A and B are straightforward accounts of pleasure, A and B themselves are more plausible when reconfigured as weakened, Aristotelian explanations (not accounts) of pleasure. I distinguished accounts/explanations that I relied on throughout the paper: Owen's understanding of A and B were two logoi of pleasure, that is, they were mutually exclusive, inconsistent accounts of pleasure and therefore could not both be true; however, explanations (*aitiai*) for Aristotle complemented one another and were able to explain different features of the same phenomena. Second, recalling that Aristotle thought each being had four possible explanations (*aitiai*) (*Phys.* II.3, 194b16-195a1), my (weakened and Aristotelian) reconfiguration of the A-B distinction was that A referred to the material explanation of pleasure: 'that out of which a thing comes to be' i.e., an activity (pleasure is an *energeiai*) and B referred to the final explanation 'end or that for the sake of which a thing is done', namely, to perfect *energeiai*. (I also classified

---

namely, an explanation that does not make a reference (implicit or explicit) to these desires, beliefs and intentions." (My italics). "Aristotle on Causality", The Stanford Encyclopedia of Philosophy (Spring 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2019/entries/aristotle-causality/>.

45 Whether this commits Aristotle to a hedonic theory of motivation I will not worry about in this paper since it hinges on larger discussions of Aristotle's action theory. Cf. Cynthia A. Freeland's "Aristotelian Actions" Nous 19.3 (1985): 397-414.

46 One objection to my *aitiai*-reconfiguration is that it does not clearly designate how to reconcile Owen's descriptions that generates the A-B dilemma. Whether as logoi or *aitiai*, the description of A and B does not seem reconciled. I have two replies. First, on my view, Aristotle's description of pleasure in VII and X are only consistent iff we take the *aitiai*-reconfiguration e.g., since for Aristotle explanations are complementary (*Phys.* II, 3), I take the description in VII to be about the material explanation of pleasure and X as the final explanation of pleasure. With respect to the Separability Thesis (ff. 44), I have argued in a rudimentary way that on my *aitiai*-reconfiguration, pleasure is not separable from the activity itself.

'being unhindered' the formal explanation, that is, 'the form' or 'definition of the essence' and efficient explanation 'that out of which a thing comes to be' the art of performing the activity in question – I used the example of 'decorating' in the paper). While Owen addressed the interpretive dilemma of pleasure in Aristotle, I was invested in a counterfactual solution to the interpretive dilemma, namely, a solution which showed how A and B could have been consistent if Aristotle had weakened A and B, as I suggested, and the philosophical merits of my reconfigured, Aristotelian solution to the A-B dilemma. In the end, Owen's project was making A and B consistent *qua* accounts (logoi); however, given the difficulties of the project and my investment in putting A and B together as explanations (*aitiai*), we ended up answering different questions. While the reader might propose an alternative account that Owen did not consider, such an alternative account, even if exegetically accurate, must be philosophically defensible such that it is superior to my reconfigured, weakened Aristotelian solution to the philosophical dilemma. Perhaps a more attractive solution of the A-B dilemma, both interpretive and philosophical, will involve an exegetical, not merely textual defense[47]; however, independent of the paper presented, it is a worthwhile project in and of itself to put Aristotle's four explanations (*aitiai*) (*Phys.* II.3, 194b16-195a1) to work in the context of his view of pleasure.[48]

❧

BIBLIOGRAPHY

Aristotle. The Complete Works of Aristotle. 2 volumes. Edited by Jonathan Barnes.

---

47 One such attempt is found in ibid., Harte, 309-314, who explicitly adopts Owen's construction of the A-B dilemma: "My proposal for reconciling claims that are made about pleasure in Aristotle's two treatments of pleasure is that, in *NE* vii 11–14, Aristotle makes only one of the two points he makes about pleasure in *NE* x 4–5. In both, he identifies as source or location of pleasure certain perfect/unimpeded activities. In *NE* x 4, but not in *NE* vii 11–14, he goes on to characterize the enjoyment component of such a pleasure as something that itself perfects an activity in a distinct, though apparently inevitable, manner." (312). It is not clear, however, whether Harte's reconciliation is successful. First, if pleasure is the source of an activity, and it is also the perfection of an activity, it is not clear what Harte's answer to the heterogeneity problem will be since Aristotle rejects the language of identity in the former and latter. Aristotle uses identity-laden language e.g., 'is' (eimi) – not deflated language e.g., source or location (*NE*, VII; *EE*, IV, 1153a9-15). Second, Harte's use of "source" is ambiguous e.g., whether it invokes Aristotle's sense of material explanations (*Phys.* II, 8), and if it does what to make of the other explanations in *Phys.* II, 8.

48 E.g., David Bostock requires, though does not himself use, Aristotle's notion of material explanation. Cf. "Pleasure and Activity in Aristotle's Ethics" Phronesis 33.3 (1988): 270.

Princeton: Princeton University Press, 1984.

Aygun, Omer. The Middle Included: Logos in Aristotle. Evanston: Northwestern University Press, 2017.

Bostock, David. "Pleasure and Activity in Aristotle's Ethics" Phronesis 33.3 (1988): 251-272.

Falcon, Andrea. "Aristotle on Causality", The Stanford Encyclopedia of Philosophy (Spring 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2019/entries/aristotle-causality/>.

Freeland, Cynthia A. "Aristotelian Actions" Nous 19.3 (1985): 397-414.

Gerd van Riel's "Aristotle's Definition of Pleasure: A Refutation of the Platonic Account" Ancient Philosophy 20.1 (2007): 119-138.

Gonzalez, Francisco J. "Aristotle on Pleasure and Perfection" Phronesis 36.2 (1991): 141-159.

Harte, Verity. "The Nichomachean Ethics on Pleasure" in The Cambridge Companion to Aristotle's Nicomachean Ethics. Ed. Ronald Polansky. New York: Cambridge University Press, 2014.

Henry, Devin. "Aristotle on Pleasure and the Worst Form of Akrasia" Ethical Theory and Moral

Practice 5.3 (2002): 255-270.

Kerferd, G.B. Review of Godo Lieberg's Die Lehre von der Lust in den Ethiken des Aristoteles. The Classical Review 10.2 (1960): 118-120

Kraut, Richard. "Aristotle's Ethics", The Stanford Encyclopedia of Philosophy (Summer 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2018 entries/aristotle-ethics/>.

Lucash, Frank. "More Pleasure in Aristotle" Rivista di Filosofia Neo-Scholastica 66.1 (1974): 126-130.

Owen, G.E.L. "Aristotelian Pleasures" Proceedings of the Aristotelian Society, 72 (1971-72): 135-152.

Plotinus. The Enneads. Edited by Lloyd Gerson. Cambridge: Cambridge University Press, 2019.

Rorty, Amelie Oksenberg. "The Place of Pleasure in Aristotle's Ethics" Mind 83.332 (1974): 481-497.

Matthew Strohl's "Aristotle on the Heterogeneity of Pleasure" in Pleasure: A History. Ed. Lisa Shapiro. Oxford: Oxford University Press, 2018, 42-65.

Munteanu, Dana LaCourse. "Aristotle on Pleasure and Learning" in Encyclopedia of the Sciences of Learning. Ed. Norbert M. Seel. Boston: Springer, 2012.

Webb, Philip. "The Relative Dating of the Accounts of Pleasure in Aristotle's Ethics" Phronesis 22.3 (1977): 235-262.

ABBREVIATIONS OF ANCIENT WORKS

Aristotle

'Cat'- Categories	'*NE*'- Nicomachean Ethics
'*EE*'- Eudemian Ethics	'*Phys. — Phys.*cs
'*An. Pr.*'- Prior Analytics	'Top' - Topics

TIME AND ETERNITY
ON "TIMELINESS" AS A LINKING MEASURE

by Everet Smith

Emory University

ABSTRACT ‖ *By engaging with two Platonic texts—the Philebus and the Timaeus—the following paper explores questions regarding the metaphysical categories that constitute time in the ancient tradition: aiôn, chronos, and kairos. In the Timaeus, Plato claims that the world is given an "aiônic" or eternal character by virtue of the perfect image upon which it is modeled. According to Heleen Keizer the use of aiôn here has long troubled scholars, as it was often thought that the aiônic quality could only be applicable to the perfect model, and not to the imperfect material copy. In response to these debates in the literature, I suggest that there is a manner in which the aiônic, or eternal, relates itself to time: the aiônic is related by virtue of kairos. Often understood as a "timeliness" or a necessity, I argue that there is an additional property of kairos that can be understood as "measurement" or metrios. I pursue an argument for the ontological or metaphysical equivalence of metrios with kairos, challenging some assumptions concerning the relationship between aiôn, chronos, and kairos; for example, that kairos is an externality that suddenly interrupts the already established relationship between aiôn and chronos. Proposing this ontological equivalency contributes to ongoing debates concerning the metaphysics of temporality insofar as "timeliness" or "the right time"*

is altered on a metaphysical scale if it is established that timeliness is immanent to the coherent whole of aiôn and chronos. By virtue of this relationship I suggest that it is always, or eternally, the right time to undertake the meaningful action that is implied in the common use of kairos. The relationship I establish between aiôn and kairos also establishes a coherent link between linear time (chronos) and eternity (aiôn), which is significant due to the former often having been considered a degradation of the latter.

The Greek term *aiôn*, commonly translated into English as "eternity,"[1] holds a distinguished position in Plato's philosophy, as many have noted.[2] However as questions concerning the nature and meaning of temporality persist still today,[3] it is crucial that we continue to consider what time is as a metaphysical category. Questions regarding structures of time or temporality are bound as well to questions concerning the nature of eternity, and the link between these two seemingly distinct temporal categories has been salient in the literature for quite some time. For example, in the Timaeus, Plato asserted that chronological time is but a moving image of eternity. The persistent question of the relationship between linear time and eternity, *chronos* and *aiôn*, is the subject I address in this paper. I argue that the nature of their relation stems from metrios—measure or mean—as expressed by Plato in Philebus as well as in the Statesman.[4] In these dialogues metrios is taken to be synonymous with *kairos*, or "timeliness." Insofar as it has been suggested by ancient theorists that *chronos*, or "linear time," emerges by way of a juxtaposition or reflection of *aiôn*, establishing *kairos qua metrios* as a linking measure demonstrates an intimate relationship between eternity, linear or chronological time,

---

1   *Aiôn*, or eternity, is rendered in Greek as "αιών." It can have multiple meanings and connotations, including "life," "life force," "life time," "whole life," in addition to its (perhaps more familiar) meaning, "eternity." For a more in-depth discussion of the history and use of this term, refer to "'Eternity' Revisited: A Study of the Greek Word αιών" by Heleen M. Keizer in Philosophia Reformata (2000).

2   Indeed, there is a pleathora of literature on the history, use, and meaning of the term *aiôn*, from its early use in Greek philosophy to its uptake in Christian doctrine (W. Kneale, "Time and Eternity in Theology," Arist. Soc. Proc., Vol. LXI, 1960-1). Keizer's study of the term *aiôn* (2010) is a valuable resource for understanding its significance in ancient Greece. A more recent example of scholarly engagements with *aiôn* includes Michel Weber's chapter in Time and Science (In 3 volumes), "The Present's Specificity," which pursues "the three Greek modes of presence (*chronos*, aion, and *kairos*)" (2023, 173).

3   For example, the recent theoretical endeavors concerning the "afterlives" of slavery, such as has been popularized by the likes of Hortense Spillers, Christina Sharpe, and others.

4   Notably, the relationship between metrios, *chronos*, and *aiôn* is affirmed later by Plotinus. See especially Enneads V1.8.

and timeliness.

The relationship between eternity, linear time, and timeliness is a metaphysical one (or, if one likes, an ontological one), as the structures of time in a first-order sense have direct bearing on the nature of one's becoming as a human being. As a metaphysical category and an aspect of what one might call first principles, how we understand these structures of time has direct bearing on our second-order sense of certain aspects of our experience—for example, ethical, aesthetic, or epistemological principles. In other words, our understanding of metaphysical substrata (primary or "first-order" principles), such as the structures of time, impacts the way we understand aspects of material life which relate to or are conditioned by these metaphysical substrata—understood here purely in the sense of primary or first order principles. While chronological time, eternity, and timeliness can indeed be considered from the perspective of the problem areas of the ethical and aesthetic, the epistemological, or even the socio-political, such implications are beyond the scope of this paper. Instead, I address the problem of the metaphysical nature of the structures of time.

In the first section, I introduce Plato's use of *aiôn* in the Timaeus, particularly as he appears to juxtapose eternity with *chronos*. Those familiar with ancient Greek philosophy may share a common understanding of Platonic philosophy, namely that it rests on a dichotomy: the ideal is opposed to the material, form to matter, the model is distinguished from the copy. Indeed, many have struggled to make sense of Plato's metaphysics, and the relationship between the Platonic Forms and the material world that they oversee. For example, William Kneale located a contradiction in the theological notion of eternity, arguing that this "self-defeating notion" in Christian theology in fact arose from an ill-fated "attempt to combine the notion of life with that of the timeless existence of the Platonic Forms," which he traces to Parmenides and Plato.[5] In the philosophy of Gilles Deleuze—particularly the work that preceded his partnership with Félix Guattari, such as Difference and Repetition (1968) and The Logic of Sense (1969)—likewise assumed Platonism posed a fundamental distinction between ideal forms and imperfect matter. (Hence Deleuze's project to "overturn Platonism," which consists in the abolition of this supposed dichotomy.[6])

---

5   M. Kneale, "Eternity and Sempiternity" in Proceedings of the Aristotelian Society, 1968-1969, New Series, Vol. 69 (Oxford University Press, 1968-1969), p. 223. See also the chapter "Aion, Khronos" (Alliez, 2004). Alliez states that while Kronos "presents us with all the characteristics of 'time,' *aiôn*, by way of contrast, is a term for which there is no modern equivalent." Alliez continues, stating that in the Timaeus, Plato "relates *aiôn* to divine life rather than to a human portion" (emphasis mine). The oppositional language deployed by Alliez is symptomatic of the understanding of Plato drawing an opposition between the model, *aiôn*, and its copy, kronos or time.

6   For further discussion see Gilles Deleuze, particularly his statement in Difference and Repetition that "the task of modern philosophy has been defined: to overturn

Amidst the abundant commentary on Plato there exists a persistent, trenchant interpretation that is commonly accepted, namely that Plato's work rests on a fundamental dichotomy. This dichotomy is in turns referred to by the names of form and matter, ideal and real, model and copy, divine and human, the eternal and time. This opposition in its specifically temporal register—that is, regarding eternity and time—is highlighted especially well by interpretations of Plato that rely on his description of the world's creation from the perfect model in Timaeus 37d. Following arguments made by Heleen Keizer, I note some theoretical difficulties encountered by those who have investigated this particular portion of the Timaeus. While *aiôn* initially appears to express an ideal quality that could only be the possession of the model, it is in fact also a quality bestowed upon the material world. Scrutinizing key passages in the Timaeus, we see that Plato does indeed grant to the material or imperfect copy the quality of the ideal. Thus, we may say that an *aiôn*ic quality is bestowed upon the chronic temporality that operates in the material world and serves as a "moving image" of eternity.

Yet, it remains an open question as to how this relationship is itself established, according to Plato. In order to address how it is that *aiôn*—a quality typically understood as operative solely in the ideal or intelligible realm—comes to be bestowed upon the material or sensible realm, I offer a reading of a passage in Plato's Philebus wherein *kairos* is positioned as the means by which *aiôn* and *chronos* come into relation with one another. However, as I show, in the Philebus this does not occur in a direct way; rather, it happens by virtue of an equivalence that Plato draws between *kairos* (timeliness) with *metrios*, measure or mean. The final section discusses how *kairos* has been understood as rupturing the everlasting Being of eternity and the chronological development of time. While this framing is compelling, a more nuanced perspective on *kairos* demonstrates that it can rather be understood as a dynamic link between *aiôn* and *chronos* by virtue of its established equivalency with *metrios* in the Philebus.

In this way, *Kairos* is established as a mediating force that links *aiôn* and *chronos*. This relationship between the different orders of time is offered by Plato both in the Philebus as well as in the Statesman. If we consider *kairos* as that which maintains a relationship between linear time and eternity, this demonstrates that, contrary to some interpretations, the ideal (eternity) and material (chronological time) are not contraries. Rather, they constitute a metaphysical whole that *kairos* mediates. Insofar as *kairos* maintains

---

Platonism" (59). Elsewhere in Difference and Repetition, Deleuze says that "overturning Platonism, then, means denying the primacy of original over copy, of model over image, glorifying the reign of simulacra and reflections" (66). See also "Plato and the Simulacrum," trans. Rosalind Krauss (1983). For further clarification and interpretations of Deleuze's anti-Platonism see Daniel W. Smith, "The concept of the simulacrum: Deleuze and the overturning of Platonism" in Continental Philosophy Review (2006).

the relationship between linear time and eternity, our understanding of "the right time" implied by *kairos* must be augmented to accommodate an understanding of *kairos* as an ever-present aspect of the whole of life. This means that eternity, chronological time, and "the right time" all work together in a reciprocal relationship.

PLATO ON AIÔN

Plato's Timaeus is an account of the genesis of the universe and of humankind. The narrator of the dialogue, after whom the dialogue is named, also gives in his account of the creation of the universe an account of the structures of time. As Heleen Keizer has noted in her research on the subject it has been relatively common to understand Timaeus' discussion of the universe as positioning *chronos*, or linear time, in opposition to eternity, or *aiôn*, or eternity. However, such a claim is highly contestable. As Keizer notes, our present understanding of eternity as a kind of timelessness may not be entirely applicable to someone writing so long ago, as Plato did. Beyond the charge of sheer anachronism, however, the assertion that Plato opposes chronological time with eternity emerges as a matter of debate in Timaeus 37d:

> As this [model] now is in fact an everlasting (aïdion) Living Being, he set out to finish also the All around us so far as possible like that. Now the nature of the Living Being happened to be *aiôn*ic (*aiôn*ios), and it was not possible to bestow that completely on what is generated; but he thought to make an image in motion of *aiôn*, and in the very act of setting the heavens in order, he made of *aiôn*, which remains at one, an *aiôn*ic image which proceeds according to number: that which we have named time (*chronos*).[7]

The model on which the world is based has an "everlasting" quality which is not possible to bestow completely upon anything which might emerge from the model's "Living Being." *Aiôn* is a quality possessed by the perfect, ideal model, and nothing else. It is not clear, however, how this quality could be bestowed at all upon an imperfect, material world. Indeed, while Timaeus himself clearly states that "it is not possible to bestow [this quality] completely on what is generated," nonetheless the creator or demiurge is successful at generating an "image in motion" of *aiôn*. This is what we have come to call linear time, or *chronos*. Yet, if *chronos* and *aiôn* are disposed to be taken as contraries in the literature, then how is it that this *aiôn*ic quality can apply to both the ideal model as well as the material copy?

There is a tendency in the secondary literature to understand Plato's work as establishing an opposition between chronological time and eternity. The material realm is of a

---

7   Plato, Timaeus in Ancient Greek Philosophy from Thales to Aristotle, eds. Cohen, Curd, and Reeve (Indianapolis: Hackett Publishing, 2011), p. 660.

fallen nature, with *chronos* being associated with decay or imperfection. Given the established affinity between *aiôn* and *chronos* in the Timaeus, however, we should explore the possibility that such an interpretation misses the mark. This is not wholly to reject that there is an important difference between *aiôn* and *chronos*. Indeed, as Timaeus continues after the above passage, the use of verb tenses such as "was" or "will be" that suggest a coming into being are often "unthinkingly but incorrectly [applied] to everlasting being [*aiôn*]."[8] This is because while *aiôn* is everlasting, unchanging, and perfect, *chronos* always implies change due to its unfolding in a countable way—for example, in our ability to keep time. However, given the applicability of *aiôn* to both the perfect model on which the universe is based, and the copy of that model that the sensible realm itself is, to insist that *aiôn* and *chronos* be understood as contraries seems not entirely accurate.

According to Keizer, Plato's use of the term *aiôn* possesses a dual connotation. On the intelligible level it connotes the unitary whole of life-time, while on the sensible level it references the display of or "counting out" of this unitary whole.[9] This is what is meant when Timaeus explains that "*aiôn* remains at one," while "an *aiôn*ic image," or the material world, "proceeds according to number." In other words, the perfection of *aiôn* may only be expressed through its relationship with the sensible, wherein the sensible realm manifests in sequence rather than remaining unified and at rest. Despite this "counting out" being associated solely with *chronos*, the above passage offers a way to understand this "counting out" as an aspect of the *aiôn*ic, precisely because it has been made of nothing other than the *aiôn*ic. Thus, another way to conceive of the relationship between linear time and eternity presents itself in this passage. Timaeus in fact attributes *aiôn* to both the ideal copy as well as the model in equal measure, with the everlasting nature of the model and the sequential, moving image of the copy differing only with regard to their respective expressions of the eternal.

There clearly must be a metaphysical relationship between linear time and eternity that goes beyond a simple opposition, for otherwise the applicability of *aiôn* to both the ideal and the material makes little sense. Timaeus 37d states that *chronos* emerges as an image of *aiôn* in motion; ergo, one might say that *chronos* emerges as an aspect of *aiôn*'s unfolding, with *aiôn* remaining eternally at rest in the intelligible realm, while *chronos* instantiates this ideal in motion in the material realm. However, insofar as each realm may claim *aiôn* as an attribute of itself and, therefore, each realm may claim perfection for itself, there must be some other way to understand this unfolding apart from presenting *chronos* as a degradation of the ideal, if not simply an abomination that constitutes an opposition to its source. I argue that this is best understood as a relation constituted by a measurement. By way of an appeal to Plato's Philebus I offer that this measurement,

8   Plato, "Timaeus", p. 661. Emphasis mine.
9   Heleen M. Keizer, "Eternity Revisited", p. 56

or metrios, is none other than *kairos*—or timeliness.

ON TIMELINESS AS A MEASUREMENT

Kairos is characterized as that which cuts across the intelligible and sensible realms in a flash, seemingly inexplicably. *Kairos* has been translated as "timeliness" or "the right time," with accompanying connotations of what is fitting, necessary, or opportune. In other words, *kairos* is described as an external disturbance that ruptures or creates a fissure in the metaphysical state of Being that constitutes eternity, and the ontic material of chronological time.

Kairos is differentiated from *chronos* by its associations with quality, not quantity. As John E. Smith notes, *chronos* connotes "the fundamental conception of time as measure, the quantity of duration," and questions pertaining to this order of time may be principally answered "in cardinal numbers or in terms of limits that approach these numbers."[10] In contrast, *kairos* is associated with the qualitative character of time, and it refers to "the special position an event or action occupies in a series, to a season when something appropriately happens that cannot happen at "any" time, but only at "that time," to a time that marks an opportunity which may not recur."[11] *Kairos*, or "the right time," marks out a special temporal position that nonetheless relates to the more ordinary progression of chronological time. Smith's characterization also emphasizes the fleeting nature of *kairos* in that "the right time" is an opportunity that can be missed.

If the opportunity is not missed—if the chance is taken at the right time—*kairos* constitutes a kind of fissure within and through *aiôn* and *chronos*, marking a space for the manifestation of an action that is necessary to initiate a qualitative change in the world. Indeed, the meaning of *kairos* has often been associated with metaphors about taking aim and firing at the right time. For example, in Euripides' The Suppliants, changing a person's mind requires "aiming [one's] bow beyond the *kairos*."[12] Illustrations of the spatial and connecting aspects of *kairos* liken it to a shuttle passing through the weaving on a loom; E.C. White, for example, describes *kairos* as the moment "when the weaver must draw the yarn through a gap that momentarily opens in the warp of the cloth being woven."[13] In both examples, *kairos* is described not only as a decisive moment for change, but also as a sort of connecting fiber between eternity and its moving image in chronological time. The weaving metaphor, in particular, holds connotations of a kind

10   John E. Smith. "Time, Times, and the 'Right Time', *Chronos* and *kairos*" in The Monist, Volume 53, Issue 1, (Winter, 1969), p. 1.

11   Ibid. Emphasis mine.

12   Euripides, The Suppliants.

13   Eric Charles White, Kaironomia: on the will-to-invent. (Cornell University Press, 1983).

of suturing of time and eternity. This suggests that a *kairon*ic moment has the ability to bring time and eternity closer together. It is this connecting aspect of *kairos* that is of particular interest to me here. If *kairos* operates as a kind of shuttle working through the fabrics of eternity and linear time, *kairos* is that force which is capable of bringing time and eternity into relation with one another. Thus, it is only through such a relation that both chronological time and eternity may share in the ideal *aiôn*ic quality, as was suggested in Timaeus.

In the Philebus, Plato discusses the tension between leading a life solely in search of physical pleasure, versus leading a life in pursuit of the Good. Philebus, a hedonist who wishes to live his life solely in pursuit of physical pleasure, is defended by the student of the Sophists, Protarchus. The latter engages in a Socratic dialogue with none other than Plato's favorite dialogic figure, Socrates, who debates with him the existence of the higher pleasures of the mind in addition to hedonist pleasures of the body. In true dialectical fashion, Socrates argues that a truly good life—a life genuinely worthy of pursuit—is one which mixes the lower pleasures of the material realm with those higher principles belonging to the intelligible realm. This supports my claim that rather than eternity and linear time being simply opposed to one another, they rather intermingle in a meaningful way. However, the means by which they come into relation or admixture is still in question. In the Philebus, the manner in which the higher and the lower pleasures enter into relation with one another is by way of what Socrates in the dialogue refers to as the chosen measure—μέτρον καὶ τὸ μέτριον καὶ (*metron* and *metrion*).

It is the principle of the chosen measure or moderation that brings the higher principles and the lower pleasures into relation with one another. Crucially, the use of Greek terminology in the concluding passages of the dialogue generates a salient equivalency between *kairos* and this 'chosen measure' of *metrion*:

> [65a] "Let us run it [the Good] down with three—beauty, proportion (symmetria), and truth (aletheia)...
>
> [66a] "Then you will proclaim everywhere, Protarchus, by messengers to the absent and by speech to those present, that pleasure is not the first of possessions, nor even the second, but first the eternal nature has chosen measure, moderation, fitness, and all which is to be considered similar to these (μέτρον καὶ τὸ μέτριον καὶ καίριον καὶ πάντα ὁπόσα χρὴ τοιαῦτα νομίζειν, τὴν ἀίδιον ᾑρῆσθαι). [14]

In the Greek terminology, here, the term καίριον or *kairion*, a version of *kairos*, emerges alongside the invocation of the "chosen measure." Indeed, one might consider here the typical translation of *kairos* as "the right time" to be another way of referring to it as a form of measurement or moderation: it is "the right time" because it is an appro-

---

14 Plato, Philebus, trans. J.B.C. Gosling (Clarendon series, 1975), p. 69-70.

priate mixture of chronological time and eternity.

Plato's equivalency between *kairos* and *metrion* in the Philebus appears again in Plato's Statesman, a dialogue aimed at articulating the figure best suited to rule justly. The figure of the statesman emerges as having a kind of specialized knowledge that makes him uniquely suited for political rule. Yet, rather than being merely in contrast to the sophist or philosopher, that statesman is best understood as the mediator between these two extremes. The equivalence of *kairos* and *metrion* emerges at 284c-e:

VISITOR (V): Is it the case then that just as with the sophist we compelled what is not into being as well as what is, when our argument escaped us down that route, so now we must compel the more and less, in their turn, to become measurable not only in relation to each other but also in relation to the coming into being of what is in due measure? For if this has not been agreed, it is certainly not possible for either the statesman or anyone else who possesses knowledge of practical subjects to acquire an undisputed existence.

> YOUNG SOCRATES (YS): Then now too we must do the same as much as we can.
>
> V: This task, Socrates, is even greater than the former one—and we remember what the length of that was. Still, it's very definitely fair to propose the following hypothesis about the subject in question.
>
> YS: What's that?
>
> V: That at some time we shall need what I referred to just now for the sort of demonstration that would be commensurate with the precise truth itself. But so far as concerns what is presently being shown, quite adequately for our immediate purposes, the argument we are using seems to me to come to our aid in magnificent fashion. Namely, we should surely suppose that it is similarly the case that all the various sorts of expertise exist, and at the same time that greater and less are measured not only in relation to each other but also in relation to the coming into being of what is in due measure. For if the latter is the case, then so is the former, and also if it is the case that the sorts of expertise exist, the other is the case too. But if one or the other is not the case, then neither of them will ever be.
>
> YS: This much is right; but what's the next move after this?
>
> V: It's clear that we would divide the art of measurement, cutting it in two in just the way we said, positing as one part of it all those sorts of expertise that measure the number, lengths, depths, breadths and speeds of things in relation to what is opposed to them, and as the other, all those that measure in relation to what is in due measure, what is fitting, the right moment, what is as it ought to be—everything that removes itself from the extremes to the middle.
>
> YS: Each of the two sections you refer to is indeed a large one, and very different from

the other.

> V: Yes, Socrates; and what many sophisticated people sometimes say, supposing themselves to be expressing something clever, to the effect that there is in fact an art of measurement relating to everything that comes into being—that's actually the very thing we have just said. For it is indeed the case, in a certain way, that all the products of the various sorts of expertise share in measurement.[15]

Again the moderate, fitting, and opportune — καιρὸν or *kairon*—is marked as equivalent to the moderate or the measured μέτριον or *metrion*. The moderate is what links eternity with *chronos*, the kind of time which we apprehend as relating to a countable, quantitative duration. *Kairos* qua *metrion* is therefore a way of comprehending how the presumed stable state of Being that constitutes eternity, as an ideal image, shifts to an image in motion.

Kairos, as a measure, marks out the temporal shift from being at rest to being in motion, and in fact binds these two modes of temporal expression together. Consider the way in which the statesman serves as the figure of the just ruler: he is characterized as a figure who exists between the extreme poles of sophist and philosopher. Drawing a comparison between Plato's characterization of the statesman and *kairos* illustrates that, for Plato, that which is capable of maintaining two, apparently opposed, extremes together in harmony is the most just. This is because holding the extremes together generates a cohesive whole. Yet, conceiving of Plato's metaphysics in this way is hardly the most common way of reading him. As my brief discussion of some relevant interpretations of the Platonic dialogues above shows, it is typical to associate Platonism with an intransigent dichotomy. This dichotomy consists in submitting what is lower to what is higher. Rather than a metaphysics based on an intransigent opposition, what Plato's dialogues time and again demonstrate is the necessity of relating two seeming opposites together, in harmony. Indeed, this understanding of Plato is one that I share with such scholars as W. von Leyden, who in the late 1960s argued that "Plato's doctrine, if examined in detail, contains considerably less contrast between the concepts of eternity and time than the traditional account makes out."[16]

Kairos is to linear time and eternity what the statesman is to sophistry and philosophy. As a just ruler, the statesman is a mediator between two extremes, holding them together, instead of existing in one realm or in the other. Rather than a fleeting chance for action that can be missed, *kairos* demands to be understood as that which brings a

---

15 Plato, Statesman in Complete Works, eds. John M Cooper and D. S Hutchinson. (Indianapolis Ind: Hackett Publishing Company, 1997).

16 Von Leyden, W. "Time, Number, and Eternity in Plato and Aristotle" in The Philosophical Quarterly vol. 14, no. 54 (1964), p. 35.

sense of wholeness to the apparently oppositional metaphysical categories making up the universe.

THREE ORDERS OF TIME AS A COHERENT WHOLE

Historically, *kairos* has been understood as a moment of timeliness or opportunity which, when manifested, creates a fissure through the ideal (eternity) and the material (time). However, upon closer inspection of key passages in Plato's Philebus, and with the support of other passages from the Statesman dialogue, it is clear that the position and import of *kairos* has heretofore been underestimated. It is indeed compelling to consider that *kairos* operates as that which bursts forth within and between the metaphysical state of Being that constitutes eternity and the ontic material of chronological time. However, *kairos* also serves as a link between *aiôn* and *chronos*, a measurement that maintains their relationship. As Plato's other dialogues reveal, this is by virtue of its established equivalency with metrios and *metron*.

Acknowledging the role of *kairos* as both a rupture and as a connection within and between eternity and linear time disrupts commonly accepted lines of argument concerned with Platonic and Neoplatonic metaphysics. For example, it suggests that rather than being positioned as mere contraries, linear time and eternity are instead intimately connected and are constantly at play with/in one another by virtue of their connection via *metrion* as *kairos*. My interpretation offers an understanding of *kairos* as being much more than just an opportunity for the proper action, which of course presents the dire possibility of missing the opportunity. If *kairos* is not only related to *chronos* as has been commonly assumed, but also to eternity, then the best way to understand *kairos* is as an ever-present opportunity that can never truly be missed, though perhaps it can be delayed for a while.

To understand my suggestion here, we ought to consider the relationship between chronological time and eternity that *kairos* establishes. Chronological time and eternity are made capable of sharing in the perfection of *aiôn* via the relational function of *kairos* as a measure. Thus, both time and eternity may be considered complete and perfect insofar as they each possess an *aiôn*ic quality. *Chronos*, therefore, no longer presents the same threat of degradation or decay in light of its relationship to the perfect whole, or *aiôn*. Indeed, this interpretation aligns with Keizer's philological research on the term *aiôn*. In "Eternity Revisited," Keizer notes:

> The Timaeus applies *aiôn* and *aiônios* exclusively and systematically on the scale of the cosmos (the model as well as the copy) as a whole. Wholeness/completeness includes here perfection. When we look at the wonderful starry image above our heads and see the heavens' rational order, movement and numbers, which is what we call time, Plato wants us to see this as the representation of *aiôn*. Interpreted in this way, time may be regarded as setting out fullness or completeness rather than duration or infinity. It is

certainly 'according to' duration, succession and even infinity that time fulfills its role, but this role consists in displaying 'life/time completeness'.[17]

Here, Keizer argues that while time qua *chronos* may operate according to a principle of duration — something that can be apprehended with "cardinal numbers," as Smith has suggested—it nonetheless is geared toward presenting life in its completeness. This is notable, considering that *aiôn* in the ancient Greek tradition, as noted in section one, apart from referring to eternity can also be translated as lifetime. As a linking unit of measure, *kairos* is in some sense ever present in its relationship with linear time and eternity. This constant presence owes to its stitching together, we might say, the fabric of time and eternity. Thus, it no longer makes much sense to say that *kairos*, as the right moment, may be missed. The "right moment" is in fact the very thing which holds together the ideal and the material, eternity and linear time.

What initially appear as three orders of time—*aiôn* as the first-order metaphysical, *chronos* as the second-order material or sensible, and *kairos* as a kind of inexplicable, spontaneous emergence—in fact constitute a coherent metaphysical whole. This whole, as suggested in the Keizer passage above, constitutes the fullness of life as time. This whole is constituted by the interplay between the three categories of eternity, time, and timeliness, which has been the focal point of my investigation here. I suggest that understanding Plato's system, and its truly dialectical implications, demands that we think these various orders of time as fundamentally and irrevocably inseparable from one another. Indeed, to think of them in any other way would appear to do each term, and perhaps our very understanding of what constitutes the world and our lives as such (that is to say, metaphysics), a disservice.

There is, however, a lingering thread that remains to be tied. While outside the scope of this paper, the argument that I have put forth here concerning the role of *kairos* as a linking unit between chronological time and eternity opens it up to having a connection with the "suddenly," or *exaiphnes*. Spyridon Rangos' work on the topic has shown *exaiphnes* is, like *kairos*, that which opens up a fissure within and between time and the eternal.[18] In this way we receive a glimpse of the eternal, and the sudden nature of this moment of rupture signals a qualitative change—again, in a manner remarkably similar to *kairos*. Referring to the Parmenides (156d-e), Ivan Platonvjak and Tone Svetelj have described *exaiphnes* as seeming "to signify that from which something changes into something else."[19] *Kairos*, Platonvjak and Svetelj note, refers to a time of transformation;

---

17 Keizer, 63.

18 Spyridon Rangos, "Plato on the Nature of the Sudden Moment, and the Asymmetry of the Second Part of the Parmenides" in Dialogue, p. 539.

19 Ivan Platonvjak and Tone Svetelj (2021). "The *Chronos* and *Kairos* of Hope" in

thus, historically *kairos* has become "the crucial rhetorical tool or moments in political, legal, ethical discussions, especially in treating emotions and mental confusion."[20] Both *exaiphnes* as well as *kairos* appear to operate as a kind of link that also severs the connection or admixture between *chronos* and *aiôn*. Yet, in bearing connotations of transformation and opportunity, they also both imply a sense of cohesion and wholeness. However, more work will have to be done to parse out the precise relationship between *kairos* and *exaiphnes*. Perhaps, like *kairos*, *exaiphnes* can function as another integral piece of the whole of time that serves to enrich both our metaphysical understanding of experience.

❧

BIBLIOGRAPHY

Alliez, E. "Aion, Khronos" in Cassin, B. (ed.) Vocabulaire Européen des Philosophies: dictionnaire des intraduisibles. Paris Seuil/Le Robert, 2004.

Cohen, Sheldon M., Patricia Curd, and C. D. C. Reeve. Readings in Ancient Greek Philosophy: From Thales to Aristotle. 4th ed. Indianapolis: Hackett Publishing, 2011.

Deleuze, Gilles, and Rosalind Krauss. "Plato and the Simulacrum." (Winter, 1983): 45-58, https://doi.org/10.2307/778495.

Euripides. "The Suppliants" in The Complete Greek Drama, edited by Whitney J. Oates and Eugene O'Neill, Jr. Random House, 1938.

Keizer, Helena Maria. Life, Time, Entirety: A Study of AiōN in Greek Literature and Philosophy, the Septuagint and Philo. [Place of publication not identified]: [publisher not identified], 1999.

——. "Eternity Revisited" in Philosophia Reformata Vol. 65, No. 1 (2000).

Lestienne, Rémy, Paul Harris, and Carlo Rovelli. Time and Science Volume 1: The Metaphysics of Time and Its Evolution. World Scientific Publishing Europe Ltd, 2023.

Leyden, W. von. 1964. "Time, Number, and Eternity in Plato and Aristotle." The Philosophical Quarterly 14, no. 54 (1964): 35–52, https://doi.org/10.2307/2955440.

Plato. Philebus, trans. J.B.C. Gosling. Clarendon series, 1975.

Plato. Complete Works. Edited by John M. Cooper and D. S. Hutchinson. Indianapolis [Indianapolis]: Hackett Publishing Company, 1997.

Platonvjak, Ivan and Tone Svetelj. "The *Chronos* and *Kairos* of Hope" in Theological Quarterly (2021), 797-806.

Rangos, Spyridon. "Plato on the Nature of the Sudden Moment, and the Asymmetry of the Second Part of the Parmenides" in Dialogue, 2014.

---

20 Platonvjak and Svetlj, 799.

Theological Quarterly, p. 801.

· SMITH ·

Smith, Daniel W. "The concept of the simulacrum: Deleuze and the overturning of
    Platonism" in Continental Philosophy Review vol 38 (2005): 89-123.

Smith, John E. "Time, Times, and the 'Right Time', *Chronos* and *kairos*" in The Monist,
    Vol. 53 Issue 1 (Winter, 1969).

Von Leyden, W. "Time, Number, and Eternity in Plato and Aristotle" in The
    Philosophical Quarterly vol. 14, no. 54 (1964): 35.

White, Eric Charles. Kaironomia: on the will-to-invent. Cornell University Press, 1983.

SOLVING THE RULE-FOLLOWING PARADOX

by LIOR TAPNACK

Durham University

ABSTRACT ‖ *A paradox is an inconsistent set of claims, each of which seems individually true.[1] To encounter a paradox is thus to find oneself caught between compelling reasons to believe contradictory things. In the case of rule-following, while people are generally confident in their ability to identify and follow rules, Wittgenstein argues that we are mistaken to feel this way. His conclusion in the Philosophical Investigations is that, concerning all possible actions, there can be no fact about one's action that indicates which rule it follows. As a result, no course of action counts as following a rule, where to follow a rule is to act in accordance with its requirements.[2] But rule-following is of central and fundamental importance to our everyday lives. When we speak, it is our ability to abide by definitions, to follow grammatical norms, that gives meaning to our expressions.[3] At a higher level, social life, law, and games are all grounded in rules too. Given the centricity of rule-following to generating meaning, this is a paradox that must be taken seriously. This paper will contend*

---

1   William Lycan, "What, Exactly, Is a Paradox?" Analysis 70, no. 4 (2010): 615–622.

2   Ludwig Wittgenstein, *Philosophical Investigations* (Oxford: Wiley-Blackwell, 2009),
    §§185–201.

3   Meaning something by a linguistic expression is analogous to following a rule. Say
    that I mean *red* by 'red'. Then my use of 'red' is correctly applied to a British phone
    box, for example, but not to a US post box. We make this distinction by referring to
    the rule for the definition of 'red'.

that there are indeed facts that demonstrate which rule an action follows – namely, intentional facts about one's dispositions to have particular introspectable mental states – and that these facts are both necessary and sufficient for rule-following. This contention, that there indeed exist facts demonstrating rule-following in action, amounts to a denial of Wittgenstein's conclusion, thereby solving the rule-following paradox.

I. MAP

Several different rule-following paradoxes have emerged out of Wittgenstein's work, perhaps as a consequence of his elusive writing style. In what follows, §2 presents the rule-following paradox as it appears in Kripke (Wittgenstein: On Rules and Private Language). §3 recapitulates on §2 but this time in logical form. §4 introduces the normativity requirement and its employment in Kripke. §5 uses the normativity requirement to refute some of the more obvious objections to the rule-following paradox. §6 introduces a more promising objection to the rule-following paradox in *sensationalism*. §7 defends this objection on its *dispositional* account, concluding with the solving of the rule-following paradox.

II. THE PARADOX

A rule is a requirement, or set of requirements, for the way something must be done. To follow a rule is thus to act in accordance with its requirements. The rule-following paradox can be derived from Kripke, though a preliminary word about the "paradox" is in order: that rule-following encounters paradox in the first instance is not universally conceded. Indeed, it is widely believed that Kripke misrepresents Wittgenstein in asserting a paradox. Boghossian's *The Rule Following Considerations* attempts to dissolve the paradox by distinguishing between rule-following as an "intentional" (as opposed to "introspectible") act, such as can be observed in mechanical algorithmic systems. Boghossian explains that following a rule must involve acting *intentionally*; one cannot follow a rule by accident. But whilst meaningful linguistic expressions must be underpinned by mental content, "on pain of regress, it cannot be that mental expressions themselves acquire meaning as a result of anyone following rules in respect of them".[4] Boghossian thereby distinguishes between "rule-following" and his preferred concept of "rule-compliance". This comports with a compatibilist and verificationist view of meaning, and thus circumventing any metaphysical considerations. Baker and Harker devote each of their three essays in *Scepticism, Rules and Language* toward arguing that skepticism

about rule-following is not paradoxical but absurd. Like Boghossian, they claim that Wittgenstein intended for the paradox to be *dissolved*, not solved, via a clarification of the relevant concepts.

Despite their harsh criticisms of Kripke's reading of Wittgenstein, Baker and Hacker hold that Kripke's rule-following considerations are profound in their own right: "while it was generally conceded that, as an interpretation of Wittgenstein's analysis of rule-following and private language, Kripke's arguments were wrong, nevertheless [...] rule scepticism [and] the rule-sceptical considerations [...] were exceedingly profound".[5] In what follows, I pay no attention to the question of Kripke's fidelity to Wittgenstein. Put otherwise, I consider Kripke's rule-following paradox as though it were his own. More importantly, I mirror Goldfarb's stance who, despite holding that the paradox in Kripke is a misrepresentation of Wittgenstein, writes that "since the (modified) physicalistic notion of fact that Kripke exploits, or something close to it, is widely shared among philosophers today, it may be worthwhile to accede to it and follow the challenge out".[6]

In this paper, I begin by presenting Kripke's argument for the rule-following paradox, which I take to be valid. A simplified version of Kripke's argument can be presented as follows:

(1) For someone to be following some rule there must be some fact in virtue of which this rule is being followed.

(2) There are no facts in virtue of which any rule can be followed.

(3) One cannot follow rules. [1, 2]

We begin by presenting the rule-following paradox as it appears in Kripke.[7] In order to set up the argument, suppose that by "+" (read *plus*) we mean the function we ordinarily refer to as "addition". That "+" means addition will be understood so long as the reader is not already convinced by the conclusion of the paradox. Suppose that I am now tasked with working out a sum with numbers larger than those I have ever calculated before. For example, let us say I am given the expression "68+57", to which I respond correctly with "125". Presumably I do so because I recognize "+" as resembling addition; by saying that 68+57 = 125, I mean that the addition function sums 125 for those numbers. The "sceptic" now asks what *meaning addition* constitutes of. For consider now an imaginary function we shall call "quus", symbolized by "⊕", where:

---

4   Paul A. Boghossian, "The Rule-Following Considerations," *Mind* 98, no. 392 (1989): 507–549.

5   G. P. Baker and P. M. S. Hacker, *Scepticism, Rules, and Language* (Oxford: Basil Blackwell, 1987), viii.

6   Warren Goldfarb, "Kripke on Wittgenstein on Rules," *The Journal of Philosophy* 82, no. 9 (1985): 476

7   Saul, Kripke, *Wittgenstein: On Rules and Private Language* (Cambridge: Harvard University Press, 1982), 7-22

$$x \oplus y \quad = \quad x + y \qquad \text{if x, y} < 57$$
$$= \quad 5 \qquad \text{otherwise.}[8]$$

Given that I have not yet added numbers greater than 56, every fact one can produce about my past actions counts equally for both plus and quus functions (for adding and "quadding"). And I cannot have been both adding and quadding at once. For what it is to add is essentially different from what it is to "quad". Eventually, the two functions output different arguments; if I were to perform both addition and quaddition in computing "68+57", I would obtain "125" and "5" respectively. In each of my responses I would be contradicting myself on account of my other response. But if I cannot be both adding and quadding at once, and if no fact I can produce about my past actions may count for either the quus or the plus function alone, so it follows that every fact one can produce about my past actions must count for neither addition nor quaddition. Herein lies the paradox. My past finite behaviors will also be compatible with many different mathematical functions beyond addition. As a result, there appears to be no fact about my past usage which says which rule I have been following, and so which answer I ought to give. Put otherwise, I can provide no *justification* for answering "125" rather than "5". For both responses can be similarly construed as the "correct" answer according to my past "rule-following" activities. Hence, neither is correct.

The concept of "quus" and "quaddition" is used here to make the problem vivid. The idea that quaddition *is* the rule I have been following is suggested merely to demonstrate my inability to claim otherwise. In other words, I *might as well* have been quadding all along. The significance of the paradox is to be found in how it generalizes. If there are no *past* facts in virtue of which I *meant* addition, then there can be no *present* facts in virtue of which I *mean* addition either. But if it can be shown that there is no fact in virtue of which *I* mean addition, then it can be shown that there are no facts in virtue of which *anyone* means addition. And if, as Arif Ahmed puts it, "there is nothing special about 'plus'", then the same can be said for any word in any language.[9] It follows that there are no facts about what *anyone* means by *anything*. And if nothing exists in virtue of which anyone means anything, then, as Kripke writes, "it seems that the entire idea of meaning vanishes into thin air".[10] For, as Goldfarb observes, "if nothing in the world settles an issue between one or another possibility, then we may conclude that there is

---

nothing to be settled".[11]

It is worth noting that the paradox of rule-following may not generalize to this degree. Above, we made a sneaky move by assuming that rule-following is a necessary component of *all* meaning. Reversing this assumption, whilst it may be true that rule-following is a necessary component of meaning "addition", there may well be other words and concepts where our inability to follow rules will not preclude our access to meaning them. Irony is one example whereby its proper use does not seem to require rule-following. Regardless, §7 attempts to solve the paradox even on its hard, "no facts" version, arguing that we can indeed follow rules.

Two further clarifications ought to be made at this juncture. Firstly, the rule-following paradox is not making an *arithmetic* claim. At no point is it arguing that 68+57 is not equal to 125. This mathematical fact and others are not in question. The paradox is also not *epistemic* in nature. It is not a debate about how one can know (with their limitations on memory) which rule one is following. Rather, the paradox is purely *metaphysical*. It is a question about what facts there *are* in virtue of which one can be said to be following rules. And, consequently, in virtue of which one can be said to be generating meaning. The paradox would persist even for the omniscient. As Kripke sometimes puts it: how can *God* know that you meant addition?

III. THE PARADOX PARAPHRASED

(1) For some person *P* to have been following some rule *R* there must be some fact *F* by virtue of which *R* has been followed by *P*.

(2) For any *F* by virtue of which *R* has been followed by *P*, *F* also instantiates some other rule contradictory to *R* being followed by P.

(3) *P* cannot have been following both *R* and its contrary.[12] [Law of Non-Contradiction]

(4) There is no *F* by virtue of which *R* can have been followed by *P*. [2, 3]

(5) If one cannot have been following a rule, one cannot continue to follow this rule.

(6) There is no *F* by virtue of which *R* can be followed by *P*. [4, 5]

(7) There is nothing special about *P*.

(8) There is no *F* by virtue of which *R* can be followed by *anyone*. [6, 7]

(9) There is nothing special about *R*.

---

8   The use of 'quus' as a new operator is sourced from Kripke and is not an invention of my own.

9   Arif Ahmed, *Saul Kripke* (London: Continuum, 2007), 103

10   Kripke, *Wittgenstein*, 22

---

11   Goldfarb, "Kripke on Wittgenstein on Rules", 474

12   I take two rules to be each other's contraries insofar as their inevitable divergence causes contradiction. For a set of rule-following acts, any two such rules are so related that one and only one can be followed. Or else, at this point of inevitable divergence, for this same set of inputs, one is both right and wrong in the application of either rule, hence infringing upon the law of excluded middle.

(10) There is no *F* by virtue of which *any rule* can be followed by anyone. [8, 9]

(11) No one can follow any rule. [1, 10]

(12) Rule-following is a necessary component of all meaning.[13]

(13) No one can mean anything. [11, 12]

IV. THE NORMATIVITY REQUIREMENT

As presented in §2, the paradox begins by casting doubt over whether any particular rule can be meant to be followed. The following criterion for meaning arises: in order to have meant something in particular there must be some fact in virtue of which I meant this particular thing. Kripke casts a further doubt, serving to constrain this criterion for rule-following. Not only must I be aware of some fact as to what I meant in the past, but present awareness of this fact must now *tell* me how to act. It must *direct* my present decision-making.[14] Kripke writes, "the 'directions' [...] that determine what I should do in each instance, must somehow be 'contained' in any candidate for the fact as to what I meant".[15] Call this the "normativity requirement" (NR), for it leads us to a set of normative truths about my behavior; obeying the NR ensures that my use of a particular expression is correct when applied to some objects and incorrect when applied to others.[16] Now, assuming that (i) what I meant by "plus" in the past is the same as what I mean by it now, and that (ii) my having meant addition by "plus" in the past is something that I am now aware of, then (iii) the NR must be in place for me to mean addition in the present.

V. OBVIOUS RESPONSES

Kripke rejects some responses (discussed below) to the paradox using the following schematic form, as highlighted by Ahmed:

(1) If F is the fact in virtue of which you meant addition by 'plus', F will satisfy the NR.

(2) F does not satisfy the NR.

---

13 Cf. the discussion regarding the relationship between rule-following and meaning in §2.

14 What kind of demand is this? Kripke does not make it clear what this 'directing' is supposed to consist in, nor does he argue clearly for employing this constraint. Since it plays an essential role in Kripke's rebuttals of the various attempts at showing what a rule-following fact consists in, one would hope for more support on this point than Kripke provides. However, we will see in §7 that adopting a 'dispositional sensationalist' position will avoid Kripke's 'skeptical solution' anyway. So I won't dwell on this point.

15 Kripke, *Wittgenstein*, 11

16 Boghossian, "The Rule-Following Considerations", 148

(3) F is not the fact in virtue of which you meant addition by 'plus'. [1, 2]

One obvious response to the paradox is to appeal to a general rule or algorithm that one is following. However, the issue here is that the paradox will rear its head at the higher algorithmic level too. Suppose I attempt to explain the addition rule by relating it to the rule for counting. I have two separate piles of items, each containing a number of items according to the sum's arguments. I merge the piles and count the total number of items in this new, larger pile. In this event, what is there to be said for my meaning counting and not "quonting," where quonting works like addition until one of the earlier piles that make up the new pile contains more than 56 items, in which case quonting outputs "5"? If I go on to explain that the concept of counting is in fact uniform for all arguments up to infinity, what is there to be said for my meaning "uniform" and not "quuniform"? And so on *ad infinitum*. Ultimately, what is there to be said about what I mean when I say anything at all? As Wittgenstein put it, the above response to the skeptic is just to give a "rule for interpreting a rule". And the inevitable regress that follows goes to show that nothing in the verbal statement of any such rule can *guide* me to say "125" rather than "5". This approach therefore fails to meet the NR.

But I do not consider myself to be choosing arbitrarily. If the assimilation of a further rule fails, I might instead claim that I knew "125" to be the correct answer all along. If I knew the correct answer in advance of the question, then I obviously meant and now mean addition. But this response fails as well. If I did indeed instruct myself in advance in this particular case, I certainly have not done so for the infinite set of larger sums. Again, this putative fact about my past actions cannot be relied upon to guide my future decisions. Therefore, it does not satisfy the NR.

Lastly, in giving the answer "125", I also cannot claim to be doing the same as I have always done. For precisely which rule I have been following thus far is the very point in question. If *doing the same* means computation according to the rule exhibited by my examples, the skeptic has already shown this to be false. As in the original formulation of the paradox, no given fact about my past actions can be said to guide my present actions. Once again, the NR is unmet and so rule-following (and generating meaning) appears to remain an impossible task.

VI. SENSATIONALISM

Perhaps the most intuitive solution to the paradox is to argue that the fact that I mean addition by "+" is (or at least supervenes on) the fact that I have some *introspectable mental state*. I have a feeling associated with "+" and it is the occurrence of this mental association by virtue of which I now mean addition and not quaddition. Let us call this view *sensationalism*. Kripke attacks sensationalism as an appropriate solution to the paradox on three counts: (1) it fails to satisfy the NR; (2) it is not *sufficient* for my meaning

addition; (3) it is not necessary for my meaning addition. Let us consider each in turn.

VII. DEFENDING SENSATIONALISM: A DISPOSITIONAL ACCOUNT

Regarding sensationalism's failure to meet the NR, Kripke puts forward two arguments. First, he observes that introspectible mental states cannot be accessed as a guide to acting in novel situations in the same way that physical models and tables can. According to Kripke, "[d]o I record and investigate the past physiology of my brain?"[17] By questioning my ability to interpret the sort of sensations or feelings that directed me in my past behavior, Kripke casts doubt over my ability to abide by the NR. For if I cannot ordinarily access these facts about my past sensations, then they cannot possibly "*tell* me how to act in the present".[18]

Second, argues Kripke, even if I could access these sensations, what is there to be said for my meaning one thing as opposed to another by them: "[n]o internal impression, with a *quale*, could possibly tell me in itself how it is to be applied in future cases".[19] As justification for this objection, Kripke references an analogy found in Wittgenstein, where the same internal impression of a cube can be shown to have more than one possible application in the physical world, depending on one's method of interpretation.[20] The resultant observation is that the relationship between my occurrent mental states and how these sensations direct me to act is far from straightforward. This leads Kripke to believe that, *even if accessible*, my sensations cannot serve to guide me in my application of a rule. This constitutes a second, distinct attack on sensationalism's ability to abide by the NR.

The first of these two objections may be true but delivers no substantial blow to the sensationalist position. Of course, one's mental states are not introspectible in the same way that a physical table is observable. As Ahmed writes, "it cannot be the case that one 'looks something up' in a mental image and then interprets it to generate an application of the associated word".[21] Remember that the paradox to which we are responding is not epistemic in nature, but metaphysical.

In response to Kripke's second objection, *how* such mental states guide our actions may well be mysterious, but *that* they guide us in some way remains intuitively clear. We even possess a basic model for this sort of sensational guidance: by way of example, the shock of pain I feel the first time I lay my hands on the coal furnace both guides me in that present moment and warns me in the future. In the case of addition and other

---

17 Kripke, *Wittgenstein*, 23

18 Ibid.

19 Ibid., 42-3

20 Wittgenstein, *Philosophical Investigations*, §139

21 Ahmed, *Saul Kripke*, 111

rule-following practices, why should we assume that the capacity for a guiding introspection no longer applies, that our mental states are suddenly void of any intentional content?

There must be some intentional content to our mental states in the application of a rule. However, the extent to which sensory phenomenology can indeed tell us something about *cognitive* functions is not totally clear. Although I experience a particular sensation when feeling pain, it is not obvious that the same type of phenomenological qualities that make up this sensation are present during my use of addition. That is, whilst facts about my sensory experiences might be sufficient to guide me in my application of a rule, perhaps there is a different kind of experience, namely my *cognitive* sensations, which are not. On this note, Kripke gives us two reasons to doubt the applicability of our learning about pleasures and pains to addition and other putative rule-following. Both reasons concern the *insufficiency* of our mental states, even if introspectable, to determine whether I am following some particular rule, whether I am actually "adding".

Note that pleasure and pain are phenomenologically distinct; one can always tell their sensory expressions apart.[22] As a result, one can appeal to, and be guided by, these mental states with ease. Perhaps this is not so with addition. What is the difference, asks Kripke, between the way it *feels* to perform the function "5+7" and the function "5×7"? Moreover, what is the difference between the way it feels to perform the function "5+7" when adding and when quadding? For each of these two objections, if there is no difference between the feeling associated with each computation, and if our feelings alone are indeed what guide us in our application of a rule, then we would have nothing within our grasp to distinguish between the correct answers to each computation. Hence, an appeal to our feelings must be insufficient for rule-following.

The first of these objections appears correct, but only for automatic computations. The sums of "5+7" and "5×7" only feel the same to me because I have learned to commit them to memory when I was a child. On the other hand, "68+57" and "68×57" would feel very different to each other. Kripke's second objection regarding quaddition, however, is immune to this response. Even sums large enough for most people not to admit of an automatic response (say "48 plus 36") should be accompanied by the same mental state as with their corresponding quums ("48 quus 36"). After all, quaddition was defined as being identical to addition for the sum of two numbers below 57; until then, quaddition just *is* addition. It follows that one will not be able to tell them apart.

---

22 Perhaps untrue: Bataille proposed that pleasure and pain can become intertwined in experiences of excess. He believed that such experiences, which he associated with the sacred, go beyond conventional distinctions and involve a "fusion of opposites", including pleasure and pain. Nevertheless, the extent to which a fusion of pleasure and pain would *inhibit* one's ability to be guided by these sensations is not inherently clear. (Georges Bataille, *Story of the Eye* (San Francisco: City Lights Books, 2001)).

This critique, however, is open to the following defense: the fact that I mean addition by "+" is the fact that I am *disposed* to give the sum of "x" and "y" as the answer to any expression of the form "x+y". On the sensationalist account, my meaning something supervenes on the fact that I am disposed to have certain sensational states in response to that thing. It follows from the dispositional account of sensationalism that what is instructive in terms of the rule I have been following is not my sensations up until this present date. Rather, it is my disposition to have the sensations associated with a particular rule in the future. With respect to the rule for addition, although there are shared mental states between addition and quaddition before reaching "68+57", what matters is not the sensations I have when performing these smaller sums. Rather, it is my disposition to answer differently once quaddition no longer traces the rule for addition (namely at "68+57" and beyond).

The following analysis of dispositions has been explicitly endorsed by Ryle, Goodman, and Quine and implicitly endorsed by countless others: an object is disposed to *M* when *C* iff it would *M* if it were the case that *C*.[23] With this analysis in mind, what makes it the case that I have been adding not quadding all along is the counterfactual claim that *had* I been asked to sum numbers greater than 57, I *would* have had the sensations associated with addition and not quaddition. We can hereby see how a dispositional account of sensationalism will circumvent a key criticism in Kripke (namely, an account of quaddition that synonymizes its effects with addition). While Kripke argues that there is no difference between the way it feels to perform the function "5+7" when adding and when quadding, the dispositionalist can distinguish between the "adder" and the "quadder" by recourse to the counterfactuals that eventually mark their distinctions.

Before I defend dispositionalism as the correct account of our sensations in rule-following, I will expand upon what is meant by dispositions from both a *linguistic* and *cognitive* perspective, in addition to the metaphysical picture already painted. This will help bridge the gap between the concept of a disposition and its manifestation in rule-following. Regarding our linguistic turn, Ryle observes that "to say that a person knows something [...] is not to say that he is at a particular moment in the process of doing or undergoing anything, but that he is able to do certain things, when the need arises".[24] Vis-à-vis rule-following, to say that somebody is following the rule for addition is just to say that he "has a certain capacity, tendency or propensity" to have that related sensation both in the present and in the future.[25]

23   Gilbert Ryle, *The Concept of Mind* (Oxford: Routledge, 2009); Nelson Goodman, *Fact, Fiction, and Forecast* (Cambridge: Harvard University Press, 1965); W. V. O. Quine, *Word and Object* (Cambridge: MIT Press, 1964).

24   Ryle, *The Concept of Mind*, 100

25   Ibid., 107

Ryle demonstrates the extent to which this much appears obvious once we reflect on the way we employ ordinary grammar: "the verbs 'know', 'possess' and 'aspire' do not behave like the verbs 'run', 'wake up' or 'tingle'; we cannot say 'he knew so and so for two minutes, then stopped and started again after a breather', 'he gradually aspired to be a bishop', or 'he is now engaged in possessing a bicycle'".[26] This constitutes Ryle's distinction between what he labels *episodes* and *dispositions* in language. In the same way that treating statements about "knowledge" and "possession" as mere episodes would be to misappropriate their semantic significance, so too would it be inappropriate to treat statements about "rule-following" in this way. A dispositional statement about someone following a rule is neither a report of observed or observable instances of a rule being followed nor a report of unobserved or unobservable instances of rule-following. Dispositional statements, Ryle claims, "narrate no incidents".[27] Instead of narrating particular actions, they tell us of a potential *to go on acting* in some way. Of course, dispositional statements are still intimately connected with narratives of incidents, for example, "Jane Doe just had the sensation for adding 5+7" *satisfies* what is asserted by the dispositional statement "Jane Doe follows the rule for addition". But what is signified by the statement "Jane Doe follows the rule for addition" is not the fact that she had the requisite sensations in the past, nor the fact that she has these sensations now, but the fact that *she will have these sensations* whenever asked for the sum of two numbers.

Ahmed clarifies any confusion about what it means to be disposed to follow a rule in terms of one's cognitive faculties. It is not that we have certain sensations that accompany our dispositions – our dispositions are not a cognitive matter – but that there are particular sensations that we are disposed to have whenever we can be said to be following a rule: "it suffices for your having meant addition by 'plus' that you were disposed to have some particular feeling in response to each query of the form 'x plus y'".[28]

But why should we feel as though a dispositional account of sensationalism (as outlined above) is an accurate depiction of our mental processes? The following thought will shed light here: *there are far too many shades of green to think of all at once*. When I say, "that is green," I may be looking at a particular green thing, but I mean green by the word green only if I am disposed to pick out certain other things as green in other contexts. Since the range of these things exceeds the sensations that I can actually have (I am, after all, finite), what makes it true that I mean "green" by the word green cannot be the sensations that I am actually having. Instead, it will be the sensations *which I am disposed to have* whenever I pick something out as green. Hence, it is only this tendency to pick out all green-like things (resp. to perform addition) that can generate meaning

26   Ibid.

27   Ibid., 108

28   Ahmed, *Saul Kripke*, 114

behind our use of the word "green" (resp. "plus").[29]

Kripke's third and final attack on the sensationalist account of meaning is to reference Wittgenstein's point regarding the dispensable nature of sensations for rule-following. One may often find themselves to be reading without any accompanying mental state of note (Wittgenstein writes of reading "out loud", where sometimes we are surprised by the accompanying sensation of reading "by heart"). This would appear to demonstrate that dispositions are not a necessary component of meaning; their instantiations (occurrent mental states) are sometimes lacking where there seems to be meaning and where rules appear to be being followed.

This is a confusing analysis of our mental states that requires some unpacking. There are three ways to explain this objection, each of which the dispositionalist can respond to without affliction. First, if the suggestion is that one really has no occurrent mental states when performing some cognitive action, this appears to be a contradiction in terms. Whether a person can appropriately be said to be acting *unconsciously* regarding a distinctly cognitive function is a matter for the biological sciences, but it certainly appears *prima facie* implausible. We may assume that either a person is not performing a cognitive action or a person possesses some occurrent mental states (but not both). If instead Wittgenstein's suggestion is that there *are* occurrent mental states but that I just do not notice them, then this is both plausible and non-problematic. The numbness of my limb is not dependent on my noticing it. Rather, the associated mental phenomena (just as with reading, addition, and so on) come into sight upon closer inspection of them. In case this does not seem obvious, consider the following: were somebody to prod me in a spot where I am numb before this numbness has come to my attention, the physical sensation I experience as a result of the prodding will manifest according to my numb state. My numbness possesses qualities independent of my immediate perceptual awareness. This allows it to persist until I direct my attention to it. Lastly, if what is meant by Wittgenstein is that we can have multiple different introspectable mental states relating to the same action – that we can have several different sensations that each accord with the rule for reading – this is fine too. After all, there are far too many shades of green to think of all at once.

CONCLUSION

I began by presenting the rule-following paradox as it appears in Kripke. I then paraphrased the argument in an attempt to show its formal validity. With recourse to

---

29   Hume, who believed dispositional sensationalism was its most intuitive form, is the only named sensationalist in Kripke's text: "[a]fter we have acquired a custom of this kind [(addition)] the hearing of that name revives the idea of one of those objects, and makes the imagination conceive it with all its particular circumstances and proportions". (David Hume, *A Treatise of Human Nature* (London: Penguin Books, 1969), I.i.7).

Kripke's normativity requirement for rule-following, I went on to reject some of the more obvious responses to the paradox, before motivating the sensationalist position as the basis upon which the rule-following paradox can indeed be solved. Therein, I defended the intuitive view that one's introspectable mental states can be appealed to as evidence for a rule being followed against a number of skeptical responses to this position in Kripke's work. I motivated a dispositional account of sensationalism in reply to the most forceful of Kripke's objections, showing that if my representation of the rule-following paradox is a fair one, then the paradox can be solved on a dispositional sensationalist account of meaning.

❦

BIBLIOGRAPHY

Ahmed, Arif. *Saul Kripke*. London: Continuum, 2007.

Baker, G. P., and P. M. S. Hacker. *Scepticism, Rules, and Language*. Oxford: Basil Blackwell, 1987.

Bataille, Georges. *Story of the Eye*. San Francisco: City Lights Books, 2001. (Originally published in 1928).

Boghossian, Paul A. "The Rule-Following Considerations." *Mind* 98, no. 392 (1989): 507–549. http://www.jstor.org/stable/2255039.

Goldfarb, Warren. "Kripke on Wittgenstein on Rules." *The Journal of Philosophy* 82, no. 9 (1985): 471–488. https://doi.org/10.2307/2026277.

Goodman, Nelson. *Fact, Fiction, and Forecast*. Cambridge: Harvard University Press, 1965. (Originally published in 1954).

Hume, David. *A Treatise of Human Nature*. London: Penguin Books, 1969. (Originally published in 1739).

Kripke, Saul A. *Wittgenstein: On Rules and Private Language*. Cambridge: Harvard University Press, 1982.

Lycan, William. "What, Exactly, Is a Paradox?" *Analysis* 70, no. 4 (2010): 615–622. https://doi.org/10.1093/analys/anq069.

McGinn, Marie. *The Routledge Guidebook to Wittgenstein's 'Philosophical Investigations'*. Abingdon: Routledge, 2013.

Quine, W. V. O. *Word and Object*. Cambridge: MIT Press, 1964. (Originally published in 1960).

Ryle, Gilbert. *The Concept of Mind*. Oxford: Routledge, 2009. (Originally published in 1949).

Wittgenstein, Ludwig. *Philosophical Investigations*, translated by G. E. M. Anscombe, P. M. S. Hacker, and Joachim Schulte. Rev. 4th edition. Oxford: Wiley-Blackwell, 2009. (Originally published in 1953).

ACKNOWLEDGMENTS