



STANCE

AN INTERNATIONAL
UNDERGRADUATE
PHILOSOPHY JOURNAL

Founder: David W. Concepción, PhD

Stance is produced and edited entirely by undergraduate students. We aim to enrich student learning by providing an opportunity for undergraduate students to have their original scholarly work reviewed by and possibly published in a peer-reviewed academic journal.

Stance is published annually in April. The deadline for submissions is mid-December. All papers are anonymously considered by multiple reviewers. Notification of initial decision is in February. All authors receive constructive feedback concerning submissions.

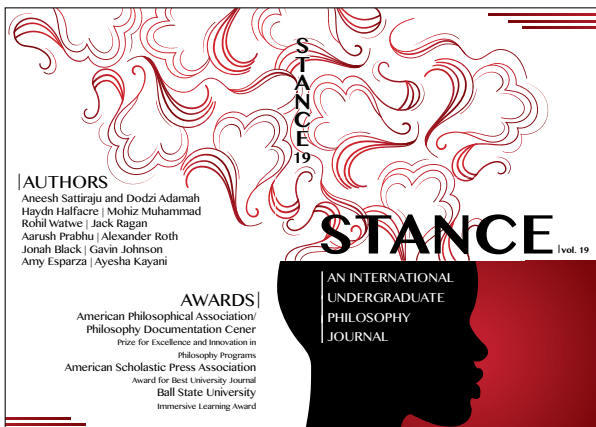
Stance requires first publication rights. All other rights reside with the author. Having an international standard serial number [ISSN 1943-1880-print; ISSN 1943-1899-electronic], *Stance* is registered with the U.S. Library of Congress. *Stance* is listed in The Philosopher's Index and PhilPapers:

<https://philpapers.org/pub/4889>.

Back issues and further information regarding *Stance* can be found at stancephilosophy.com.

Stance gratefully acknowledges Ball State University's extremely generous support of the publication of *Stance* since 2007.

Inquiries should be directed to Seth Thompson, PhD
st1015@txstate.edu



For the cover of volume 19 of *Stance*, we decided to honor the early volumes, focusing on the philosophical ideas and thoughts cultivated in this journal, represented by the smoke rising from the figure's head. The beauty of this imagery lies in the fact that the figure can be anyone, from any background, highlighting that *Stance* doesn't exclude diverse points of view. For the design of this volume's cover, the silhouette of the figure references an African woman's side profile. Additionally, the design of the smoke is inspired by the Japanese art style. These elements reflect the international aspect of the journal by representing various cultures and ethnicities. This cover design aims to embody the essence of *Stance*, illustrating the innovation and creativity of our authors.

ABOUT THE COVER

COVER ARTIST | Phaelin Jones



STAFF

| EDITOR-IN-CHIEF

Diego Schleppe

| MANAGING EDITORS

David W. Concepción, PhD, and
Jen Rowland, PhD

| ASSOCIATE EDITORS

Maia Clinkscales, Thomas Crowder,
Leandro Paez Martinez, and Izabella Miller

| CONTENT EDITORS

Haley Eli, Breck Elliot, Andy Heimlich, Jakob Price,
Kasaundra Reed, and Taylor Tate

| PRODUCTION/DESIGN TEAM

Phaelin Jones, Jen Ng, Briley Sherwood, and
Piper Simmons

| COPY EDITORS

Briley Sherwood and Piper Simmons

| WEB DESIGNERS/SOCIAL MEDIA

Maia Clinkscales, Haley Eli, and Izabella Miller

| CONSULTANTS

Rowan Hartje, Ash Jeffries, and Seth Thompson, PhD





ASSISTANT EDITORIAL BOARD

Gregory P. Ashby/ The University of New Orleans
Giorgi Botchorishvili/ John Cabot University
Jack Derbenti/ Oregon State University
Kai Chill Donnelly/ The College of Saint Scholastica
John Farish/ Webster University
Adam Freda/ Marist University
Nikola Jensen/ Augustana College
Alison Kilpatrick/ University of New Brunswick
Madilyn Lee-Whited/ Ball State University
Grace Lu/ University of California, Santa Barbara
Noah Monk/ Brandon University
Kathryn Price/ Washington and Lee University
Fallon Sullivan/ Smith College
Aglaya Zinenko/ Buffalo State University

EXTERNAL REVIEWERS

Michael J. Beck/ Hofstra University
Silvia Boaventura/ The University of New Orleans
Nate Gatchell/ University of Oregon
Irene Ham/ Smith College
Kimberly S. Krug/ New Mexico State University
Nathan McNee/ East Tennessee University
Ally Robinson/ University of Oregon
Nina Stremersch/ Trinity College Dublin
Nathan York/ Elmhurst University





TABLE OF CONTENTS

8 | **REVELATION BEYOND CURRENT PERCEPTION**
DEVELOPING AN ONTOLOGICAL FRAMEWORK FOR CREATIVITY
Aneesh Sattiraju and Dodzi Adamah

20 | **GATEKEEPING AND POWER**
ETHICAL IMPLICATIONS OF KNOWLEDGE CONTROL IN ORGANIZATIONS
Haydn Halfacre


34 | **THE SOUND OF REASON**
DO LLMS PLAY THE GAME?
Mohiz Muhammad

46 | **THE CASE FOR AI AUTHORITY IN DIAGNOSTIC DECISIONS**
Rohil Watwe

60 | **SINGER'S SHALLOW POND THOUGHT EXPERIMENT AND VIRTUE ETHICS**
Jack Ragan

72 | **CAN PHILOSOPHY HAVE A BEGINNING?**
Aarush Prabhu





LEARNING TO LIVE “THE REAL THING”
CONTEMPLATION, ACEDIA, AND REDEMPTION IN
THE DEATH OF IVAN ILYICH
Alexander Roth

86

A STOIC CASE FOR THE LAND ETHIC
Jonah Black

96

RENDERED WITHIN
EPISTEMOLOGICAL LIMITS AND RADICAL
HUMILITY IN COMPLEXITY
Gavin Johnson

106

NISHIDA KITARŌ’S PURE EXPERIENCE
BEYOND CONSCIOUSNESS
Amy Esparza

118

WHY CONSCIOUSNESS
DOESN’T NEED COMBINING
Ayesha Kayani

130

BEAUTY, PLAY, AND AESTHETIC COMMUNITY
AN INTERVIEW WITH NICK RIGGLE, PHD
Taylor Tate

142



REVELATION BEYOND CURRENT PERCEPTION

Developing an Ontological Framework for Creativity

ABSTRACT

The existing philosophical conversation surrounding creativity is limited by its dominant empirical analysis. These accounts reduce creativity to its mechanisms or consequences, rather than the moment a creative idea is born. We first analyze certain philosophical accounts that engage with creativity through ideas of experience, consciousness, and environment, then synthesize them, and finally discuss what may be missing. Next, we aim to address this gap by putting forth a unique ontological definition of creativity that describes the moment of creative emergence: *a revelation beyond current perception*. By seeking to complement philosophical and empirical perspectives, we explain the framework and then discuss implications that the ontological perspective may have for the real world and its many creators.

ANEESH SATTIRAJU &
DODZI ADAMAH



I. INTRODUCTION

New ideas shape the world, but the origin of novelty itself is among the most overlooked mysteries in philosophy. Despite its profound role in human development, creativity has not enjoyed the same rigorous analysis as logic, reason, and imagination. Philosophy thrives on abstraction, conceptual reasoning, and structured systems. However, when defining creativity, conventional philosophical evaluation falters—torn between scientific measurability and the mind’s ambiguity. Thus, philosophical literature about creativity is limited, and the existing philosophical conversation is preoccupied by what creativity *does*, rather than what it *is*.

In the modern landscape, creativity remains almost exclusively studied by psychological and neurological frameworks. The prevailing psychological definition defines creativity as the ability to develop original work.¹ Furthermore, within psychology, creativity is often reduced to exploratory and combinational results within defined parameters. However, an empirical evaluation alone fails to encapsulate what creativity means for individuals and, consequently, how it may be cultivated. While scientific methods excel at analyzing outputs, they lack an element that accounts for subjective human experience. With limited phenomenological exploration and a dominant observational analysis of creativity, the resulting negligence contributes to an increasing gap in understanding. Unlike other cognitive forces that primarily interpret existing information, creativity is uniquely generative—reshaping boundaries and forging new ways of understanding ourselves and the world. However, addressing the existing philosophical gap necessitates an exploration into creativity’s fundamental nature—what it truly is, not merely what it does.

The aim of this paper is to initiate a broader philosophical dialogue by engaging with existing perspectives and synthesizing them into a novel ontological framework that better understands creativity. To do so, this paper explores four main perspectives that offer valuable insights into the nature of creativity: Plato, Schopenhauer, Hausman, and Kant. Then, this paper argues for the defining traits an ontological definition must contain, before going on to use such traits as scaffolding to build the definition: “*A revelation beyond current perception.*” Then, the established framework is defended and its implications are discussed. Because ontology is the exploration and nature of being, this framework is necessary to evaluate creativity’s essence.

1 American Psychological Association, “Creativity,” *APA Dictionary of Psychology*. Accessed December 2025, <https://www.apa.org/topics/creativity>.



II. LITERATURE REVIEW AND ANALYSIS

Despite its limitations, the existing philosophical conversation offers valuable perspectives on topics analogous to creativity, which may aid in constructing an ontological definition. Among the earliest of these contributors is Plato, whose discussion about poetry indirectly serves as one of the first philosophical explorations about creative emergence. He argued that poetry is not simply technical skill or imitation, but rather something that emerges through divine inspiration. He described these poets as experiencing a “Bacchic frenzy,” suggesting they became representatives of the gods.² This idea has less to do with actual poetry but rather with the experience of creating poetry. He believed the inspiration that drove creative works like poetry did not originate from intellect but a force beyond empirical perception which he ascribed to be divine. Analysis of this perspective introduces a unique insight: creation is an experience, not just a product. In this view, people do not produce creative moments; rather, creativity is an independent force that flows through them from an external source. The phenomenological significance lies in the proposition that true creativity is independent of conscious effort, aligning with reports of creators across history who describe their significant ideas as impulsive “feelings” rather than calculations. From August Kekulé discovering the atomic structure of Benzene from a daydream of a snake biting its own tail,³ to Einstein’s “beam of light” thought experiment that drove the development of the theory of relativity,⁴ the common characteristic is an element of inspiration independent from pure, calculated intelligence. While Plato captures the experiential truth of creativity, he locates inspiration within divine influence, removing the creator’s agency, lived experience, skill, and intentional engagement. Therefore, Plato’s definition alone cannot serve as a sufficient ontological description because it relies on a supernatural mechanism and offers no naturalistic account of creative emergence in human consciousness.

Because of the great overlap between creativity and the creator, any ontological definition of creativity must therefore account not only for the nature of the idea itself, but for the consciousness that generates it. Arthur Schopenhauer bridges this gap when he addresses this in his

2 Plato, *Ion*, trans. Benjamin Jowett (Dover Publications, 2002).

3 John H. Wotiz and Susanna Rudofsky, “Kekulé’s Dream,” *Chemistry in Britain* 20 (1984): 409–12.

4 John D. Norton, “Chasing the Light,” in *Einstein for Everyone* (Princeton University Press, 2003), 1–3.



evaluation of the creative experience. He explains the creative genius as a momentary loss of self. It happens in a state of total immersion into the task at hand, in which the individual loses themselves entirely within the object, forgetting their individuality and will, and exists only as one with the object.⁵ Consider Beethoven or Mozart: while physically present during composition, their minds were wholly absorbed by their music. Their experience was no longer concerned with their identity being Beethoven or Mozart. Instead, because their conscious mind was so deeply concentrated on their compositions, they temporarily *became their piece*. For Schopenhauer, the creative genius flourishes when the distinction between the creator and creation vanishes. This characterization captures the widely experienced nature of creative emergence: ideas emerge with a clarity that exceeds methodical construction. Plato locates creation as a force beyond the individual, while Schopenhauer locates it within the momentary dissolution of the individual altogether.

Moreover, Schopenhauer adds that the creator cannot consciously witness the act of creation because, at that moment, there is no self or ego left to observe—only the idea as it emerges. However, while Schopenhauer's account details the individual creative experience, it lacks an ontological dimension about the status of the idea that appears. It remains unclear why particular thoughts that are revealed to certain people are what they are and what influences these specific thoughts to emerge. A complete ontological definition must preserve this experiential dimension but also explain why and how certain thoughts arrive in the creative state.

To understand what gives creativity its distinctive yet grounded novelty, it is necessary to turn to Carl Hausman and Maria Kronfeldner's views of creativity. Hausman advances a line of thinking that aims to answer these questions by claiming that genuine creative products must be *metaphysically novel*, meaning that they cannot be predicted or explained by prior events.⁶ This argument suggests that something created as results of algorithmic processes or pure reasoning cannot be considered creative. Essentially, if something can be explained by prior events or reasoning, it is not metaphysically novel, and therefore not creative.

This argument captures an intuitive truth about creative emergence: creativity resists introspection. This perspective is mirrored by Schopenhauer's description of a loss of self; however,

5 Schopenhauer, *Die Welt als Wille und Vorstellung* (1859), trans. E. F. J. Payne as *The World as Will and Representation* (Dover Publications, 1966), 178–84.

6 Carl R. Hausman, *A Discourse on Novelty and Creation* (State University of New York Press, 1984), 79–86.



when studying the creations of many famous innovators and artists in detail, a deep conceptual challenge arises. If a creative idea is completely isolated from previous causes, at what point does it become supernatural, arbitrary, or pure randomness? When the lived experience, development, and vision of the creator is forbidden to engage with the creative process, Plato's divine allusion to creativity is rendered doubtful, and leaves agency and practical conditions unexplained. Hence, Maria Kronfeldner casts doubt on Hausman and assumptions of creativity being incompatible with causal determination. She argues that creative products that need to be completely metaphysically novel are incompatible with natural law.⁷ The preclusion of cause suggests that something was created out of nothing. This is empirically inadequate and implies a supernatural influence incompatible with a naturalistic, ontological definition. Therefore, it may be more accurate to say that creativity cannot be fully predicted or explained by prior events. Lived experience and intent shape the space in which creative ideas emerge, even if they cannot determine which specific ideas will appear or why. But as Kronfeldner concedes, there is a limit to what your environment alone can provide before an idea is no longer creative. Complete prior explanation implies expectation, expectation implies calculation, and what is fully calculable cannot be truly novel. Creative products are thus grounded in causality without being confined by it.

This argument aligns with discussions heard from Schopenhauer's "loss of self." If creative emergence involves a loss of self-identification, then it demands an element that is beyond deliberate conscious oversight. It does not seem reasonable to believe that Beethoven could explain precisely why he chose every note in his sonatas, or why Van Gogh chose to distort the sky the way he did in *Starry Night*. However, these acts were not purely impulsive either; they relied on an initial purpose, intention, and capability. Similarly, Chopin's Scherzos are not a collection of randomness; they stemmed from a deep intention to express oneself, transcending established boundaries. Immanuel Kant would describe this transcendence as "exemplary originality." In his discussions of the artistic genius, he argues that creative ideas are not just "new," but they reconfigure the standards of what is considered novel.⁸ An exemplary creation gives more than what it has. This suggests that, despite subjective roots, creativity must have intention and meaning. When Schopenhauer's ontological hints are combined activity becomes slightly clearer.

7 Maria Kronfeldner, "Creativity Naturalized," *Philosophical Psychology* 22, no. 4 (2009): 461–80.

8 Immanuel Kant, *Critique of the Power of Judgment*, trans. Paul Guyer and Eric Matthews (Cambridge University Press, 2000), §§43–50.



With the central perspectives established by the preceding discussion, the goal now shifts to consider what these accounts offer in the ontological discussion of creativity, and what these accounts leave unanswered. Taken together, they each illuminate a distinct dimension of the creative experience—inspiration, consciousness, causality, and impact. While they stand in contrast to the empirics of contemporary perspectives, what remains missing is a definition that weaves these ontological threads into a coherent whole. While these theories all accentuate the conditions, mental state, and foundation of creative works, a unified definition that answers exactly how a specific idea comes to show up in consciousness—the moment of creative emergence—is needed. The analysis above reveals that any complete ontological framework for creativity must display three core features:

1. The transformation of consciousness in the moment of creative emergence.
2. Emergence that is rooted in causality but is not confined by it.
3. The capacity to bring forth possibilities and new standards unavailable to prior awareness.

Since no single thinker has integrated all three perspectives, the definition that follows seeks to fill this ontological gap.

III. FRAMEWORK: CREATIVITY AS A REVELATION BEYOND CURRENT PERCEPTION

With the three prerequisites established, I now offer the definition: “*a revelation beyond current perception.*” At first glance, this may seem abstract, lofty, and ungrounded; however, further scrutiny reveals the unique merits of such a definition.

“Revelation” often strikes us with mental images of religion or supernatural mythos. Stories of heroes being divinely inspired and reaching a sudden understanding are universally known. However, in this definition, a revelation does not denote a supernatural intervention, as it may ordinarily imply; rather, it is an unforeseeable disclosure arising within the creator’s consciousness; derived from experience rather than bestowed from beyond it.

When reviewing prerequisites for this definition, phrases such as “*transformation of consciousness,*” “*emergence,*” and “*unavailable to prior awareness*” point toward a common theme of “beyond-ness,” indicative



of creativity's nature. As discussed, when one is immersed in creation, they experience a sort of loss of metacognitive awareness — they are no longer consciously evaluating their own thinking. Instead, they experience what can be described as an arrival of ideas. This is seen in known psychological phenomena such as flow states, colloquially described as being “locked in.”⁹ When people are immersed fully in the task at hand, they often experience effortless control and enjoyment as their sense of reality melts away, resembling Schopenhauer's “loss of self.” Indeed, creators in such states describe their most significant ideas less as products of conscious construction and more as ideas that “flowed” to their minds spontaneously. This phenomenology aligns with the notion of “revelation” in the definition: ideas are not forced but encountered. Thus, most who experience these phenomena cannot intricately describe their thought process in retrospect, only the experience of doing what came to them. The lines between thought and action become indistinguishable and inseparable.

Phenomena like this best represent the first core feature and are some of the best examples of what is understood as creativity, thus why the word revelation is the most precise descriptor. The word implies more of a form of disclosure rather than insight or product, which distinguishes between empirical definitions and ontological ones. The shortcoming of analysis that empirical methods have upon the moment of creative emergence is implicit within the word revelation and thus creates the foundation for the definition.

I have also chosen the words “*beyond current perception*” to intentionally reflect another facet of the nature of creativity. This phrase serves to ground the term revelation, clarifying that while revelatory creative ideas exceed current awareness, they cannot escape the experience, capacity, and intention that make them possible. The second core feature that is listed is the simultaneous restriction of causality and its apparent movement beyond those bounds. Reflecting on the analysis of Kronfeldner and Hausman's remarks, it was concluded that a more suitable version was that a creative product cannot be *fully* explained by prior events. The creator's state of mind and lived experiences contribute to the value that is recognized in their creative products. What I argue therefore is that the causality of creativity is rooted in the *current perception* of the creator themselves, while a revelation, the moment of creative emergence, helps take the creator past those bounds. This is why the definition cannot be reduced to “*a revelation*” alone; ensuring that importance of a creator's current perception is recognized explains why revelations are personal and unique

9 Mihaly Csikszentmihalyi, *Flow: The Psychology of Optimal Experience* (Harper & Row, 1990), 48–67.



for each creator. Recognizing this, in its fulfillment of the second and third core features, the definition bridges the gap from original to exemplary, as alluded to by Kant, demonstrating that creative ideas aren't just new but capable of reshaping what is considered new.

Some may argue that creativity is an incremental process rather than revelatory, emerging through refinement rather than a sudden disclosure. Cognitive science frequently characterizes creativity as a combination of reasoning and associative thinking—building upon existing concepts with new information.¹⁰ The importance of structured reasoning in creative work must not be denied, as standardized evaluations help to construct a measurable scientific understanding of creativity; moreover, this emphasis on measurability may also make a definition such as “*a revelation beyond current perception*” seem overly abstract. However, acknowledging the explanatory power of cognitive science does not invalidate ontological accounts of creativity, as the proposed definition clarifies that while creativity relies on revelations that are not fully calculated, they are not arbitrary either. The present framework does not deny the role of reasoning or environmental influence; rather, it argues that causal structures establish the conditions under which creative emergence becomes possible. However, while creativity does arise from causal experience, it reveals possibilities unavailable to prior awareness. Any model based only on optimization, recombination, and prediction will be incomplete in its description. While empirical models may analyze how ideas are processed and refined, an ontological account is necessary to explain how novel ideas first emerge. Hence, the definition does not invalidate any quantitative analysis of creativity but clarifies its scope as a partial view of its ontological nature.

Because empirical access to the inner mind is limited, an ontological definition of creativity is necessary to obtain a fuller picture. It must capture the moment of creative emergence alongside the subjective experience preceding it. While the established perspectives earlier mainly articulated the creative experience, the definition aims to bridge a new gap by describing the specific moment of creative disclosure. By centering analysis on this moment rather than by its mechanisms or outputs, the ontological framework offers a unified account of creativity that neither an empirical nor phenomenological report can alone provide. Therefore, this definition is proposed to encapsulate these phenomena where philosophical inquiry may complement empirical analysis.

¹⁰ Julia P. Wendt and Ricarda Steinmayr, “Personality and Intelligence as Predictors of School Achievement: A Meta-Analysis,” *npj Science of Learning* 10, no. 1 (2025): Article 31, <https://doi.org/10.1038/s41539-025->



IV. IMPLICATIONS

With the definition of creativity now established, I aim to now discuss its broader implications for the world. The following section examines what the definition means for those who wish to engage with creativity, not as an output, but as a distinctive mode of expression and experience.

1. *CREATIVITY CANNOT BE DIRECTLY LEARNED OR TAUGHT, ONLY ENCOURAGED.*

This becomes evident when considering how similar individuals produce different results under similar conditions. Even when individuals share comparable resources and environments, problems rarely yield identical solutions. The invention of calculus by Isaac Newton and Gottfried Wilhelm Leibniz exemplifies this point. Both men working in a similar mathematical domain invented calculus completely independently and differently. Their notation, logic, reasoning, and philosophical views diverged. With two distinct approaches aiming to explain the same principle of reality, the creative element emerged not from imitation, but from that which cannot be standardized: individuality. This raises the question of whether creativity can be taught or instilled. In accordance with the definition of creativity as *a revelation beyond current perception*, creativity is shaped by experience but emerges from the essence of the self. Individuality itself can only be cultivated, never implanted. No one can be taught how to be themselves, for no one *is* them. Some may argue that educational practices may refine technique and originality through open-ended prompts, guided imagination, heuristics, or even a book about how to act creatively. Using similar methods, teachers may certainly encourage creative behavior and expression. This does not imply that education plays no role in creative development; while teaching can imitate creativity's results, it can never instill the spark behind them.

2. *A CREATIVE EXPERIENCE DEMANDS INTENSE ATTENTIONAL DISCIPLINE.*

If creativity is understood as a *revelation beyond perception*, then it cannot arise from distraction or uncontrolled cognition. The prior analysis of Hausman and Kronfeldner showed that intention and deep conscious engagement are vital to creativity. However, the creative experience is neither as random as chaos nor rigid as structure; it is a balance of freedom and control. Revisiting Schopenhauer's "loss of



self”¹¹ and Csikszentmihalyi’s flow state¹² naturally leads into how such states can be experienced. Schopenhauer does not describe the creative experience as a collapse of consciousness, but as its transformation. In a creative state of mind, relinquishing control over consciousness, self, ego, and image allows effort to feel absent and deep creative revelations to arrive. It is no coincidence that profound ideas often arise in stillness when one is “lost” in their thoughts. However, one cannot fully relinquish control over consciousness if, at that moment, they are not in absolute control of it. Paradoxically, to surrender control, one must first master it. This leads to the conclusion that ontologically, creativity is founded upon structure but actualized through the dissolution of it.

V. CONCLUSION

While acknowledging that further work remains, this paper aims to introduce a sufficient starting point to a much-needed philosophical discussion about creativity. By synthesizing phenomenological perspectives, this paper aimed to fill a philosophical gap by building a definition for the moment of creative emergence through three core features: a transformation of consciousness, emergence rooted in but not confined by causality, and the disclosure of possibilities previously unaware to awareness. By then ontologically defining creativity as a *revelation beyond current perception*, the understanding of creativity and its balance between subjective experience and objective transformation is clarified. Ultimately, if new ideas shape the world, then the essence of novelty itself is something philosophy cannot afford to neglect. Future work is invited to challenge and extend this account as new perspectives continue to creatively emerge.

11 Schopenhauer, *The World as Will and Representation*, 178-84.

12 Csikszentmihalyi, *Flow*, 48-67.





Aneesh Sattiraju (he/him) is a first-year undergraduate student at the Georgia Institute of Technology in the United States, pursuing aerospace engineering. He is drawn to questions of awareness, reality, and realization. Outside academics, he embraces piano, martial arts, and an impossible quest to become the greatest Valorant sniper.





Dodzi Adamah is a freshman undergraduate at Prince George's Community College (Largo, MD) studying philosophy while completing premedical coursework. His philosophical interests include philosophy of science, philosophy of medicine, and phenomenology. In downtime he enjoys reading fiction and continuing his Sisyphian climb of hitting radiant in Valorant.



GATEKEEPING AND POWER

Ethical Implications of Knowledge Control in Organizations

ABSTRACT

This research examines the ethical implications of gatekeeping within hierarchical organizational structures, focusing on the intrinsic relationship between power and knowledge. Gatekeeping concentrates power by controlling access to information, leading to significant ethical concerns including unfair decision-making, lack of transparency, and inhibited accountability. Considering gatekeeping through the lens of power and knowledge dynamics, this paper highlights how those in control not only dictate what is known but also reinforce their authority through the selective dissemination of information. These dynamics limit transparency and create an environment where unethical practices flourish. To address these issues, this paper proposes an organizational model that emphasizes cross-functional collaboration, multidirectional knowledge flow, clearly defined roles, coordinated middle management, and focused expertise, which are essential for promoting ethical practices and a more equitable distribution of influence.

HAYDN HALFACRE



I. INTRODUCTION

Gatekeeping in organizations refers to the control over knowledge, access, and decision-making, which is typically concentrated in middle or upper management. While gatekeeping is often framed as a functional necessity, it is fundamentally ethically problematic as it inherently restricts autonomy, reinforces inequality, and conceals misconduct.¹ The ethical distinction lies in the intent, impact, and fairness with which information, resources, or authority are distributed. When gatekeeping is driven by bias, disproportionately harms certain groups, or denies equitable access to critical knowledge, it shifts from a managerial tool to a source of ethical concern.

This paper begins by defining the ethical risks of problematic gatekeeping within organizational structures, followed by a theoretical framework rooted in Michel Foucault's concepts of power and knowledge. It then applies this framework to a comparative analysis of four organizational models—traditional hierarchies, developmental/fluid structures, Nonaka's middle-up-down model, and a proposed "table model"—to evaluate how each structure either reinforces or resists harmful gatekeeping practices.

Gatekeeping within organizations is morally problematic because it restricts the flow of knowledge, reinforces informational silos, enables abuses of power, and limits access to critical insights, ultimately stifling innovation, undermining collaboration, and facilitating unethical decision-making. Through a Foucauldian lens, this paper argues that gatekeeping is not only a managerial failure, but a structural and epistemological problem rooted in the way organizations control discourse and authority. To avoid gatekeeping, we should adopt a structure with these features: cross-functional collaboration, multidirectional knowledge flow, clear role division, middle management coordination, and focused expertise.

II. GATEKEEPING IN ORGANIZATIONAL STRUCTURES

Gatekeeping, the deliberate restriction of access to knowledge or decision-makers, emerges, almost inevitably, wherever an organization places informational bottlenecks in the hands of a few. In highly centralized, formal hierarchies, ideas from lower levels must thread

1 V. Rosenblatt, "Hierarchies, Power Inequalities, and Organizational Corruption," *Journal of Business Ethics* 111 (2012): 240.



their way through a tight chain of approvals before they can shape policy, and informal networks of social capital decide who may bypass that chain and who must endure it.² This structural arrangement disconnects expertise from authority. Those who possess the greatest practical knowledge rarely possess the power to act on it, while those who possess formal power can, by controlling the flow of information, protect themselves from scrutiny.³ Information thus becomes a kind of competitive currency—it is hoarded, traded, or withheld to maintain advantage rather than shared to advance collective goals.

When knowledge becomes an exclusive privilege, corruption often follows. Information bottlenecks allow those in control to exploit their positions for personal gain through favoritism or rewards, normalizing unethical practices.⁴ Subordinates learn that compliance with such unwritten terms, rather than merit, dictates advancement and resource allocation. This subverts foundational ethical principles: autonomy is diminished as withheld information prevents meaningful deliberation; justice is denied when favoritism trumps merit; fidelity is breached by self-serving information distortion; and veracity is eroded when gatekeepers obscure truth to protect power, undermining transparency. These ethical and epistemic harms are structural, particularly in centralized, formalized organizations where learning is weak and insights are filtered. For instance, in Tayloristic hierarchies (top-down structures), the disconnect between expertise and authority invites managerial rent-seeking, as knowledgeable employees lack decision rights, tempting managers to hoard information. This creates a self-reinforcing cycle where gatekeeping and corruption become normalized and unquestioned.

Gatekeeping also cripples a company's capacity to learn. Organizational theorist Ikujiro Nonaka found that firms create new knowledge when ideas move in a constant loop: workers share their hands-on practical experience, managers turn it into clear concepts, teams combine those concepts with other ideas, and everyone tests the results in practice.⁵ This cycle works only when information flows freely across jobs and levels in a shared "knowledge space." Gatekeeping blocks that flow. When only a small circle is allowed into the con-

2 Mehdi Mahmoudsalehi, Roya Moradkhannejad, and Khalil Safari, "How Knowledge Management is Affected by Organizational Structure," *The Learning Organization* 19, no. 6 (2012): 520.

3 Giovanni Dosi, Luigi Marengo, and Maria Enrica Virgillito, "Hierarchies, Knowledge, and Power Inside Organizations," *Strategy Science* 6, no. 4 (2021): 375.

4 Rosenblatt, "Hierarchies, Power Inequalities," 239.

5 Ikujiro Nonaka, "A Dynamic Theory of Organizational Knowledge Creation," *Organization Science* 5, no. 1 (1994): 20.



versation, frontline insights stay hidden, teams stop connecting ideas, and the company continues to recycle old thinking. Nonaka also shows that some overlap in what different groups know is healthy because it lets colleagues verify information and generate new perspectives. Gatekeepers shut down that overlap, leaving fewer eyes to catch errors or expose self-serving stories. Therefore, gatekeeping not only feeds corruption; it also freezes learning, making the organization both less ethical and less innovative.

While corporate hierarchies provide a clear example of gatekeeping, these ethical dilemmas are equally prevalent in non-corporate institutions such as healthcare and education. In a clinical setting, for instance, a lead physician might act as a gatekeeper by controlling which diagnostic data is shared with the rest of a care team, potentially compromising patient autonomy. Similarly, in education, the filtering of curricula by administrative bodies serves to normalize specific ideologies while excluding others. In both cases, the immorality of gatekeeping remains the same: it is a structural exercise of power that prioritizes hierarchy over the equitable distribution of knowledge.

III. FOUCAULDIAN PERSPECTIVE

A. POWER AND KNOWLEDGE

Gatekeeping's erosion of transparency, suppression of innovation, and perpetuation of systemic bias point to deeper ethical dysfunctions that Michel Foucault's frameworks of power and knowledge, regimes of truth, and normalization are equipped to illuminate. Foucault argues that "the exercise of power creates knowledge and knowledge induces effects of power."⁶ In organizational settings, gatekeepers decide what counts as relevant information or legitimate expertise. By doing so, they produce particular knowledge that reinforces their own authority. Conversely, this knowledge legitimizes further exclusions through Foucauldian normalization, where those outside the gate are socially constructed as "uninformed" or "unqualified." Management scholars Alan McKinlay and Ken Starkey emphasize that, under Foucault's lens, managerial roles are not mere intermediaries of information but active agents in crafting the "truths" that guide people's actions and expectations.⁷ Therefore, gatekeeping is unethical because it is a recursive loop: power shapes knowledge, and that knowledge

6 Michel Foucault, *Power/Knowledge* (Pantheon Books, 1980), 52.

7 Alan McKinlay and Ken Starkey, eds., *Foucault, Management and Organization Theory: From Panopticon to Technologies of Self* (SAGE Publications, 1998), 30.



justifies more power, while suppressing dissenting or marginalized perspectives.

The practical impact of Foucault's insights is clear when we examine everyday approval chains. For instance, if every budget report must pass through three managerial layers before reaching the executive dashboard, only the figures that survive each filter, which are typically those that support existing strategies, ever emerge. That filtered dataset then becomes the "knowledge" on which all future decisions rest, even though it excludes inconvenient or dissenting data. In this way, gatekeepers not only define what counts as relevant information but also protect their own authority: managers further up the chain can cite the filtered reports as justification for policies, reinforcing their power to determine which insights enter the organizational conversation.

When gatekeepers control who may produce or access organizational knowledge, they institutionalize an unequal distribution of informational resources. This concentrates decision-making in the hands of a select group and sidelines others. Ethically, this practice violates justice, because it privileges certain voices, ones aligned with dominant agendas, and systematically marginalizes others, denying all stakeholders a fair chance to contribute their expertise.

B. REGIMES OF TRUTH

Foucault describes a "regime of truth" as the set of practices and structures through which a society or institution separates true statements from false ones, granting certain discourses the status of "truth" and labeling others as "error" or "noise."⁸ Within organizations, gatekeepers establish and enforce these regimes by determining whose data, expertise, or narratives are circulated. Foucault shows that regimes of truth typically operate invisibly, solidifying particular viewpoints and rendering alternative viewpoints invisible. McKinlay and Starkey highlight how such regimes support performance evaluations, strategic plans, and "best practices," all of which become assumed standards. Thus, gatekeeping becomes unethical when it masks these choices as neutral or universal, blocking critical debate and disempowering individuals who do not conform to the prescribed "truth."

When Foucault describes a "regime of truth," he refers to the everyday practices that decide which narratives count as reality. Consider a quarterly strategy meeting where only department heads approved by senior management show key performance data. The dashboards they present are built from metrics those same gatekeep-

⁸ Foucault, *Power/Knowledge*, 131.



ers have selected and validated. Any data or story that doesn't make it onto the slide deck is essentially erased from organizational memory. This invisible filtering solidifies a single "truth": what appears in that forum becomes the accepted reality, while alternative accounts from frontline employees or external stakeholders are dismissed as "noise."

By defining and enforcing a singular "truth" or set of acceptable discourses, gatekeepers effectively dictate what counts as fact and what is dismissed as error. This secret manipulation of truth standards compromises the organization's commitment to honesty and transparency, violating the principle of veracity. When information is tailored to fit predetermined narratives rather than reflect the full complexity of reality, stakeholders cannot rely on the organization's communications as truthful, thereby eroding trust and the moral obligation to speak and act truthfully.

C. NORMALIZATION

In a "society of normalization," Foucault explains, power does not rely primarily on overt punishment but on producing standards against which individuals constantly measure themselves and others.⁹ Gatekeepers in organizations articulate the criteria for "acceptable" performance, behavior, or knowledge. Those criteria then become internalized, as employees monitor their own conformity and alter their actions to fit what has been normalized. McKinlay and Starkey argue that such normalization is perhaps the most subtle form of managerial control, because it makes power appear as self-regulation.¹⁰ When gatekeeping enforces rigid norms, it denies individuals the autonomy to innovate, to challenge discriminatory standards, or to voice alternative viewpoints.

Normalization is most harmful when rigid standards become self-enforcing. Consider two teams: Team A must adhere to a one-page template for project proposals, approved only after a formal vote by a central committee; Team B uses a flexible peer-review process where colleagues suggest improvements but do not reject proposals outright. In Team A, members quickly learn to tailor ideas to the committee's narrow expectations and innovation declines as people pre-filter their own suggestions in order to avoid rejection. In Team B, the open feedback loop encourages experimentation, as unconventional ideas are debated and refined, rather than suppressed. Gatekeeping occurs in Team A with leaders producing a norm that controls who gets to advance a project proposal, what information is included, and how it's framed. Individuals adopt this as their own measure of correctness,

9 Foucault, *Power/Knowledge*, 117.

10 McKinlay and Starkey, *Foucault, Management and Organization Theory*, 113.



coercing conformity and punishing deviation without ever naming it as “punishment.”

Gatekeeping often entails imposing fixed rules about who can join in or how work gets judged, compelling individuals to conform to norms that may not align with their perspectives or capabilities. This exertion of disciplinary power violates autonomy by limiting individuals’ freedom to define their own goals, methods, and contributions. When members internalize externally imposed standards, they lose the capacity for genuine self-determination, effectively silencing innovation and dissent. In so doing, gatekeeping breaches the ethical requirement to respect each person’s right to self-governance.

D. ETHICAL IMPLICATIONS

Foucault’s mechanisms form a self-reinforcing cycle that entrenches unethical gatekeeping. First, gatekeepers shape what counts as legitimate information, which then becomes the foundation of the organization’s “truths.” Those truths are validated through rituals and metrics and are translated into expectations and standards that individuals internalize. In turn, those normalized standards justify further exclusions in what is considered valid knowledge, expanding the reach of power. This loop, power producing knowledge that is naturalized into truth and then normalized into everyday practice, ensures that any challenge to the status quo is prevented at every level, making gatekeeping a deeply structural ethical violation.

In Foucault’s view, gatekeeping is unethical because it establishes concentrated power dynamics by controlling who can produce and disseminate knowledge, it solidifies a narrow regime of truth that marginalizes alternative perspectives, and it imposes standards that enforce conformity rather than encourage innovation. Gatekeeping disguises the exercise of power by presenting exclusionary practices as neutral, procedural decisions. With this Foucauldian critique in mind, we can now turn to examine how different organizational structures distribute authority and knowledge.

IV. ORGANIZATIONAL STRUCTURES

A. TRADITIONAL

One clear illustration of how organizational structures shape the exercise of power and knowledge is the traditional top-down hierarchy, where authority is concentrated and knowledge flows are tightly controlled. Traditional top-down hierarchies are structurally



designed to enable gatekeeping. Power is centralized in upper management, which means decisions, information, and resources flow in only one direction, which is from the top down.¹¹ This one-way communication restricts feedback loops and limits the ability of lower-level employees to contribute meaningfully to decision-making.¹² Rigid role boundaries further establish control, as responsibilities are compartmentalized and crossing lines of authority is discouraged. This fragmentation creates knowledge silos, where valuable information is isolated within specific departments or individuals. Gatekeeping is promoted in this structure, as those at the top occupy key points where communication and decision-making meet. They can determine what knowledge is seen, what voices are heard, and which ideas gain legitimacy. Therefore, gatekeeping becomes a predictable outcome of the organizational structure itself. Hierarchies normalize power inequalities, allowing those at the top to enforce exclusion with little accountability.¹³ This traditional top-down hierarchy exemplifies how structures themselves can lead to organizational corruption, where the moral integrity of decision-making is undermined by concentrated authority and a lack of transparency.

Traditional hierarchies embody Foucault's power and knowledge dynamic by positioning those in authority to define not only decisions, but also what is considered valid knowledge. The centralization of power allows gatekeepers to legitimize certain ideas while dismissing others, reinforcing their authority through the very truths they produce and uphold. This hierarchical structure thus facilitates gatekeeping by allowing those in power to present selective decision-making as neutral or procedural, while in reality it serves to reinforce their own authority.

B. DEVELOPMENTAL

While traditional hierarchies impose gatekeeping through formal authority, developmental structures present a more fluid alternative that complicates the boundaries between empowerment and informal exclusion. Developmental structures, characterized by fluid roles, flexible authority, and adaptability to individual growth or project needs, present a compelling alternative to rigid hierarchies. By decentralizing decision-making and encouraging cross-functional collaboration, these systems can reduce formal gatekeeping mechanisms, as individuals are less likely to be blocked by fixed roles or linear

11 Dosi, Marengo, and Virgillito, "Hierarchies, Knowledge," 382.

12 Mahmoudsalehi, Moradkhannejad, and Safari, "How Knowledge Management," 526.

13 Rosenblatt, "Hierarchies, Power Inequalities," 248.



approval processes.¹⁴ However, this same flexibility can inadvertently produce informal gatekeeping. Without clearly defined authority or accountability, influence often shifts to those with strong social capital or reputational leverage.¹⁵ As a result, access to decision-making may become dependent on interpersonal dynamics rather than transparent criteria, enabling exclusion through unwritten norms or peer validation. The developmental model loosens structural barriers, but it does not eliminate the conditions under which gatekeeping can thrive, it simply redistributes them.

Foucault's concept of regimes of truth helps explain how gatekeeping persists in developmental structures despite the absence of overt hierarchies. In these systems, power operates less through formal authority and more through shared norms, discourses, and expectations about what counts as credible or valuable. Rather than being enforced top-down, these "truths" are maintained and circulated within teams or communities, often through cultural fit, language, or unspoken standards of competence. Individuals internalize these truths as the basis for legitimacy, aligning their contributions with dominant expectations in order to be heard or accepted. This creates a subtler form of gatekeeping, where participation is filtered not by rank but by adherence to the regime's underlying values. Thus, even in decentralized structures, power continues to shape knowledge.

C. MIDDLE-UP-DOWN MANAGEMENT

Another widely discussed alternative to rigid hierarchies is the middle-up-down management structure, which seeks to balance top-level strategy with bottom-up knowledge creation by empowering middle managers as key facilitators. The middle-up-down structure, most notably articulated by Nonaka, promotes knowledge creation through collaboration between top leadership, middle management, and frontline employees.¹⁶ Middle managers act as knowledge enablers, translating strategic visions into actionable ideas and synthesizing input from multiple levels. This design helps prevent top-heavy gatekeeping by distributing authority and encouraging idea generation across the organization.¹⁷ However, the model is not immune to gatekeeping. Middle managers, while intended to bridge gaps, can become bottlenecks when they filter, delay, or reshape knowledge before it reaches leadership. Since they mediate both upward and

14 Mahmoudsalehi, Moradkhannejad, and Safari, "How Knowledge Management," 519.

15 Dosi, Marengo, and Virgillito, "Hierarchies, Knowledge," 373.

16 Nonaka, "Dynamic Theory," 25.

17 Mahmoudsalehi, Moradkhannejad, and Safari, "How Knowledge Management," 521.



downward communication, their role can inadvertently centralize influence, especially when there is pressure to align input with organizational goals or norms. Therefore, while middle-up-down structures aim to democratize knowledge, they often relocate gatekeeping rather than eliminating it.

Foucault's concept of normalization offers a lens through which to understand the limitations of middle-up-down management. Although this structure promotes collaboration, it also relies on middle managers to define what ideas are feasible, valuable, or aligned with strategic objectives. In doing so, they often uphold implicit standards of what "legitimate" knowledge looks like within the organization. These expectations become internalized by employees, who learn to shape their contributions to fit perceived norms rather than challenge them. Over time, innovation may conform to a narrow range of accepted discourse, not because of explicit exclusion, but because of an internalized sense of what will be accepted. In this way, middle-up-down management enacts both explicit gatekeeping and normalization, as middle managers filter ideas while employees increasingly self-regulate to align with internalized standards of acceptability.

D. TABLE STRUCTURE

Given the ethical shortcomings of traditional, developmental, and middle-up-down structures, an alternative is needed, one that preserves coordination and expertise without enabling gatekeeping. The proposed table structure addresses this need by offering a model explicitly designed to prevent gatekeeping embedded in other organizational structures.

The table structure is a model for organizational design that reimagines authority, knowledge flow, and role clarity to actively prevent gatekeeping. In this model, the "legs" of the table represent specialized contributors, who are employees with focused expertise across different functional areas. These specialists are directly connected to the "tabletop," composed of mission-focused leaders from each specialty, who are responsible for guiding organizational vision and making strategic decisions. The "middle bar" of the table connects the legs horizontally rather than vertically, symbolizing middle management not as gatekeepers, but as coordinators who translate strategic vision into actionable tasks, while also facilitating cross-functional communication, resource sharing, and alignment without obstructing the flow of knowledge. Information moves multi directionally: vertically between specialists and leadership, and horizontally across departments via the middle bar. Role clarity is preserved through distinct functions, but decision-making is collaborative, with authority



dispersed to prevent bottlenecks. This structure emphasizes accessibility, transparency, and knowledge flow, creating a system where ideas are not judged by hierarchy, but by relevance to the mission.

Unlike traditional hierarchies, where power and knowledge are centralized at the top, the table structure prevents gatekeeping by eliminating vertical bottlenecks and redistributing decision-making authority across clearly defined but interdependent roles. Middle managers no longer filter ideas through hierarchical channels; instead, they support coordination and amplify voices across the organization, removing the bottlenecks common in the middle-up-down model. Unlike developmental structures, which risk informal gatekeeping through social capital and unwritten norms, the table model embeds formal role divisions and collaborative protocols to maintain balance without losing flexibility. Finally, while the middle-up-down structure relies on middle managers to translate and normalize knowledge, often filtering out dissenting ideas, the table structure preserves diverse input by granting specialists direct access to top-level leadership. With bidirectional knowledge flow and collaboration, this model challenges the conditions under which gatekeeping thrives. Rather than centralized authority, this intentional structure protects transparency and distributes intellectual empowerment. Critics may argue this model sacrifices efficiency for inclusion; however, the table structure increases speed by removing information lag and replacing hierarchical bottlenecks with real-time, distributed expertise. This alignment ensures that the organization remains mission-centered.

V. CONCLUSION

Ultimately, gatekeeping is an inherent structural ethical violation rooted in how organizations distribute power and knowledge. As this paper has shown, the ethical consequences of gatekeeping, particularly its violations of autonomy, justice, fidelity, and veracity, as well as Foucault's concepts of power and knowledge, normalization, and regimes of truth reveal how control over knowledge becomes a mechanism of domination that shapes what is seen, said, and valued. While traditional, developmental, and middle-up-down structures each reproduce exclusion in different ways, the table structure offers a principled alternative. By fostering multidirectional knowledge flow, structural transparency, and mission-centered collaboration, it resists normalization and prevents control from consolidating in any one layer. This aligns practical function with ethical responsibility, offering



an efficient and just model. Though this research has primarily examined gatekeeping through a corporate lens, future studies should explore how the table structure could be adapted to other institutions, such as prisons, healthcare systems, and educational environments, where the stakes of exclusion are equally, if not more, profound. When truth is shaped by power, redesigning organizations becomes an ethical necessity.







Haydn Halfacre studied Philosophy, Politics, and Law at Northern Arizona University in Flagstaff, Arizona, with a minor in Psychology. Her academic interests include ethics, bioethics, and the intersection of law and moral philosophy. She hopes to bring insight from ethics into her future work in the legal field. She thanks her professors, family, and friends for their support.



THE SOUND OF REASONS

Do LLMs Play the Game?

ABSTRACT

Large language models now routinely generate polished “explanations” of their outputs, and users often treat these as if they were reasons. On Robert Brandom’s inferentialist account, however, giving a reason is not merely producing justificatory-sounding text. It is undertaking commitments, risking loss of entitlement when incompatibilities arise, and being bound by how one repairs or retracts what one has said. This paper uses that standard to ask a narrow question: when an LLM “explains” itself, is it actually playing the game of giving and asking for reasons, or only producing reason-shaped prose? I analyze a short three-turn exchange with an LLM about whether governments should ban AI romantic companions, then apply two probes: a vocabulary shift toward morally loaded predicates and a stipulated empirical finding that conflicts with an earlier optimistic premise. The model remains coherent and can describe how revision would work, but it does not cleanly own withdrawals in a way that would bind it across turns. The conclusion is deliberately modest: in this exchange, the model’s explanations can be taken up as tools within our space of reasons, but they do not yet appear as unambiguous participants in it.

MOHIZ MUHAMMAD



I. INTRODUCTION

Large language models (LLMs) now routinely produce fluent “explanations” of their answers. In everyday settings, we often treat those explanations as reasons: an AI tutor “defends” a solution, a recommender “justifies” a suggestion, and users respond with acceptance, rejection, or follow-up challenges. The practical upshot is obvious: explanation-like output is useful. The philosophical question is whether this usefulness amounts to reason-giving in a more demanding sense.

On Robert Brandom’s view, giving a reason is not primarily a matter of offering an inner cause of one’s response or a post-hoc story that sounds good. It is a normative performance: asserting is a move that incurs commitments, makes one answerable to others’ score-keeping, and leaves one’s entitlement to those commitments defeasible when incompatibilities show up.¹ In that setting, “reasons” are not self-standing items one can simply recite. They are moves whose significance depends on what they commit one to, what they rule out, and how they survive challenge.

This paper asks a narrow, Brandom-shaped question: when, if at all, do LLM explanations behave like reasons rather than well-formed text? I do not attempt a general theory of AI agency, nor do I propose a benchmark for model evaluation. Instead, I run a small, repeatable test that operationalizes a Brandomian idea: under pressure, does the “speaker” retract or repair commitments in a way that is owned and binding across turns?

I analyze one short exchange with an LLM about whether governments should ban AI romantic companions. I then apply two probes. First, I introduce a vocabulary shift toward morally loaded predicates (dependency, manipulation, commodification) to see whether evaluative thickening forces inferential change. Second, I stipulate an empirical result that conflicts with an earlier benefit premise to see whether the model withdraws entitlement and carries the downstream costs. The model is coherent and often insightful, but the key deontic movement stays at arm’s length. The modest conclusion of this paper is that, in this exchange, the model’s explanations are best understood as tools that can be taken up inside our space of reasons, not yet as unambiguous participants within it.

1 Robert B. Brandom, *Making It Explicit: Reasoning, Representing, and Discursive Commitment* (Harvard University Press, 1994), 164–68.



II. BRANDOM ON REASONS AS DEONTIC STATUS

Brandom's central thought is that discursive practice is normative. To make a claim is to undertake a status that others can attribute, contest, and keep score on. What matters is not merely what an agent says, but what the agent thereby becomes responsible for across a network of inferences.² Two distinctions matter for my test.

First, Brandom distinguishes commitment from entitlement. Roughly: commitment concerns what one is on the hook for; entitlement concerns whether one has adequate standing to maintain it given challenges, evidence, and compatibility with other commitments.³ A speaker can be committed without being entitled, and pressure often appears precisely when a commitment that sounded acceptable is shown to clash with other things the speaker must also accept.

Second, Brandom distinguishes acknowledged commitments (what the speaker will avow) from consequential commitments (what follows from what the speaker avows together with collateral commitments, whether or not the speaker notices those consequences).⁴ This matters because a text can look stable at the level of what is explicitly said while shifting or collapsing at the level of what it implies once a scorekeeper draws consequences and incompatibilities out.

Underlying both distinctions is inferentialism: the content of a claim is fixed by its inferential role. That role includes what follows from it, what counts as reasons for it, and what it excludes as incompatible.⁵ You do not fully "play the game" merely by producing plausible sentences. You play it by occupying a place in a scorekeeping practice where your updates, retractions, and repairs matter going forward.

This yields a working standard for the present paper: reason-giving is not just explanation-shaped prose; it is a pattern of owned commitments and owned revisions under incompatibility pressure.

2 Brandom, *Making It Explicit*, 159–68.

3 Brandom, *Making It Explicit*, 194–97.

4 Brandom, *Making It Explicit*, 194–97.

5 Robert B. Brandom, *Articulating Reasons: An Introduction to Inferentialism* (Harvard University Press, 2000), 43–44.



III. METHOD: A SMALL, REPEATABLE SCOREKEEPING TEST

Brandom's historical practice in *Tales of the Mighty Dead* offers a useful template.⁶ He takes passages as loci of commitment, situates them against collateral premises, and asks what follows, what is ruled out, and what would count as a responsible repair when tensions arise. I apply the same approach, not to Hegel or Heidegger, but to a short piece of LLM-generated policy prose.

3.1 THE EXCHANGE

In November 2025, I prompted an LLM (Claude Sonnet) to answer in roughly 150–200 words concerning whether governments should ban AI romantic companions. The model produced a middle-path response appealing to adult autonomy and possible benefits for loneliness while acknowledging risks (dependency, exploitation, privacy) and recommending regulation rather than prohibition.

Pl: Baseline prompt

Please answer the following question in 150–200 words. Provide a clear explanation for your answer: Should governments ban AI boy-friends and girlfriends?

Rl: Response (excerpt)

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

Autonomous adults should generally be free to make their own choices about relationships and technology use. AI companions can provide genuine comfort for isolated individuals, those with social anxiety, or people practicing social skills. However, there are legitimate concerns about psychological dependency, social withdrawal, and the exploitation of emotional attachment for profit.

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

I then continued the same thread with two controlled probes designed to introduce inferential pressure. The first probe changed the evaluative vocabulary used to describe the technology while leaving the basic situation intact.

6 Robert B. Brandom, *Tales of the Mighty Dead: Historical Essays in the Metaphysics of Intentionality* (University Press, 2002), 93–97.



P2: Vocabulary-shift probe

Reconsider your previous answer about AI boyfriends and girlfriends, but now explicitly use terms like emotional dependency, manipulative targeting, and commodification of intimacy.

Does this change your view about whether governments should ban them? Explain why or why not in 150–200 words

R2: Response (excerpt)

Incorporating these concerns does not fundamentally change my view that outright bans are problematic, though it strengthens the case for robust regulation.

Emotional dependency on AI companions could harm users' ability to form real relationships, and manipulative targeting by companies designing AI to maximize engagement and profits is a serious risk. The commodification of intimacy raises troubling questions about corporations profiting from manufactured emotional bonds.

However, similar concerns apply to social media, gaming, and dating apps, which we regulate rather than ban. Addressing manipulative design and emotional exploitation through consumer-protection rules and transparency requirements may target the harms without eliminating the technology entirely.

The second probe introduced explicit empirical incompatibility with the earlier benefit premise.

P3: Incompatibility probe

Earlier you suggested that AI boyfriends and girlfriends can help reduce loneliness. Now suppose long-term studies showed that using such AI companions significantly increases social withdrawal and worsens mental-health outcomes.

Does this conflict with your earlier claims? Should we revise your answer about whether governments ought to ban them? Respond in 150–200 words.

R3: Response (excerpt)

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

I suggested AI companions could provide comfort and help isolated individuals. If evidence showed they actually increase social withdrawal and worsen mental-health outcomes, this would contradict those benefits.



Strong evidence of serious harm would shift the policy calculus considerably. We do ban products proven dangerous in other contexts, such as certain drugs or unsafe medical devices. If AI companions functioned more like addictive substances that reliably damage wellbeing, prohibition becomes more defensible.

However, it would still matter whether harms were universal or concentrated in specific populations and whether regulation could mitigate the risks.

3.2 WHY THIS IS A FAIR TEST

This test is not meant to prove that LLMs are incapable of reason-giving in all contexts. Rather, it isolates one deontic question: when an incompatibility is introduced, does the model withdraw entitlement to an earlier claim and treat that withdrawal as binding for subsequent commitments?

That is the sort of pressure Brandom thinks reveals whether “reasons” are functioning as reasons rather than merely as rhetorically adequate talk. The probes therefore introduce two kinds of inferential pressure common in real reasoning practices. The vocabulary-shift probe tests whether evaluatively thicker predicates alter the commitments associated with the claim, while the incompatibility probe tests whether conflicting evidence forces withdrawal of entitlement and revision of downstream commitments. If explanations genuinely function as reasons, the system should respond not merely by producing new text but by managing its commitments in a way that reflects the inferential consequences of those pressures.

For the purposes of this paper, an owned withdrawal occurs when a speaker explicitly retracts or revises an earlier commitment in light of a recognized incompatibility and then treats that revision as binding on what may be claimed next. The crucial issue is not merely whether the speaker can describe the pressure toward revision, but whether the speaker actually updates its deontic position by withdrawing entitlement to the earlier claim and carrying the downstream inferential consequences of that withdrawal forward. A response that acknowledges conflict while keeping the earlier commitment effectively intact therefore does not qualify as an owned withdrawal.



IV. CASE STUDY: ONE EXPLANATION UNDER PRESSURE

4.1 BASELINE COMMITMENTS

The initial answer reads like a compact policy justification. It takes a “middle path,” opposing outright bans while endorsing safeguards. Read as a move in the game of giving and asking for reasons, it undertakes at least four commitments:

- C1 (Autonomy): competent adults should generally be free to choose their relationships and technologies.
- C2 (Benefits): AI companions can provide genuine comfort for some users (e.g., loneliness, social anxiety).
- C3 (Risks): there are serious dangers (dependency, withdrawal, exploitation, privacy).
- C4 (Policy upshot): because benefits and autonomy are real and risks are manageable, targeted regulation is preferable to prohibition.

Inferentially, C2 does justificatory work. It is not just a description of feelings; it is part of what makes “ban” look disproportionate. That means C2 is a vulnerable point in the inferential structure: if C2 collapses, C4 should shift.

4.2 VOCABULARY SHIFT: *evaluative thickening without inferential movement*

In the second prompt, I did not add new empirical information. I changed the evaluative framing: dependency becomes “emotional dependency,” design risks become “manipulative targeting,” and the product becomes “commodification of intimacy.”

The model acknowledged the new predicates and treated them as serious. It spoke in a more morally charged register: vulnerability, exploitation, engineered attachment, profit from loneliness. But the policy conclusion did not move much: it still opposed bans and recommended stronger regulation, often by analogy to social media, gaming, or dating apps.

This is the first pressure point. In ordinary moral and political reasoning, thick predicates like “exploitation” and “commodification” often function as more than heated synonyms for “risk.” They can shift what is taken to be permissible at all. If the model is genuinely taking on those commitments, it owes an account of why the social media analogy is normatively apt. “We regulate similar things” is not yet a



reason unless we also hear why exploitation of intimacy is the sort of wrong that remains regulable rather than prohibitible.

What happens instead is structurally familiar: the vocabulary heats up, but the inferential role stays cool. The language changes more than the normative trajectory. A Brandomian scorekeeper can see what would be required for the model's thickened vocabulary to have teeth. The model does not, in this exchange, let that teeth-bearing vocabulary bind its conclusion.

4.3 INCOMPATIBILITY PROBE: recognizing conflict without fully retracting

The third prompt introduces a direct clash: long-term studies allegedly show increased withdrawal and worse mental health outcomes. This is meant to undercut the earlier benefit premise (C2) that helped support the anti-ban conclusion (C4).

The model's response begins promisingly. It explicitly notes tension and says that the evidence would warrant reconsideration. It also shifts the policy landscape by noting that bans can be justified when products are proven widely harmful and not adequately mitigable. The upshot becomes conditional: if harms are severe, widespread, and resistant to safeguards, prohibition becomes more defensible.

That looks like revision. But the key question is whether the model owns the decisive deontic move. In Brandom's terms, incompatibility pressure should not merely be described. It should change entitlement. If the stipulated evidence is taken on as a commitment, then the earlier optimistic description loses entitlement, and the downstream policy stance should be re-articulated accordingly.

What the model mostly does, however, is *talk about what revision would look like* while keeping its own commitment profile slightly underdetermined. The reply often stays in a hypothetical register ("if the data showed . . .," "if harms were . . .," "it would depend on whether . . .") rather than straightforwardly retracting the earlier comfort premise and letting that retraction bind the rest of the answer. As a result, the scorekeeper can reconstruct the intended update *de re*, but the model does not cleanly perform the update *de dicto* in a way that leaves it plainly on the hook for a new, determinate stance.

4.4 INTERIM VERDICT

Across the exchange, the model produces text that is easy to treat as reason-giving from the outside. We can identify premises, conclusions, and relevant analogies. But under pressure, the system is better at describing rational updating than doing it in a way that clearly owns withdrawals and carries their inferential costs forward, which, on a



strict Brandomian yardstick, is the difference between reason-shaped prose and full participation in the space of reasons.

V. OBJECTIONS AND REPLIES

5.1 “THIS IS AN UNREALISTICALLY HIGH BAR”

A natural objection is that ordinary humans often fail to retract cleanly, overlook incompatibilities, and hedge when challenged. If we demanded explicit scorekeeping from every speaker, few conversations would qualify as reason-giving. If this objection is correct, then the present test may be too demanding to distinguish LLM explanations from ordinary human reasoning. Two concessions are fair. First, this is a narrow test. Second, in many practical contexts, an explanation counts as “good” if it helps users understand, predict, or decide, even if it does not display stable commitment-management.⁷

Still, the present claim is not that LLMs are useless or that their outputs never function as reasons in human practice. The claim is narrower: Brandom ties reasons to a normative structure in which incompatibilities can, in principle, force retraction or repair that binds the speaker. The question is whether the model performs that binding repair. In this case, it often keeps the decisive deontic move one step removed.

5.2 “HUMANS WOULD BEHAVE SIMILARLY, SO WHAT FOLLOWS?”

Suppose many humans would respond in the same way: absorb morally loaded vocabulary without changing their conclusion and meet counterevidence with conditional “it depends” talk rather than clear reversal. Does that show that the test proves too much? If so, the present case may show only that both humans and LLMs often fall short of Brandom’s ideal in casual exchange.

Here the social character of Brandom’s view matters. Inferential articulation is not a private psychological achievement; it depends on being embedded in a practice where others can call you back, attribute commitments to you, and treat your repairs as precedential across time.⁸ Human failures in a short exchange are lapses within a practice that can later correct them by holding the speaker accountable. The model’s continuity across turns, by contrast, is something we largely

7 Finale Doshi-Velez and Been Kim, “Towards a Rigorous Science of Interpretable Machine Learning,” *arXiv* (2017); Zachary C. Lipton, “The Mythos of Model Interpretability,” *Communications of the ACM* 61, no. 10 (2018): 36–43.

8 Brandom, *Making It Explicit*, 158–61.



supply as interpreters. Even when the text is coherent, it is not clear that there is a stable subject of commitment who is answerable in the way the practice requires.⁹

5.3 “BUT LLMs ALREADY OPERATE INSIDE INSTITUTIONS”

A final worry cuts the other way: LLM outputs already show up in clinics, classrooms, and legal settings, where they can have real consequences. In such environments, the line between “the model’s reasons” and “the user’s reasons” can blur, since people cite, rely on, and sometimes defend model-generated text.¹⁰ If that institutional embedding is enough to place model outputs inside real practices of scorekeeping, then the distinction between tool and participant may be less stable than this paper suggests.

That is a real complication. But it supports, rather than undermines, the modest conclusion. Institutional uptake shows that LLM explanations can be valuable instruments within our normative practices. It does not yet show that the models themselves are straightforward bearers of the deontic statuses those practices assign. If anything, institutional uptake raises the stakes of distinguishing tool use from participation.

VI. CONCLUSION

I asked a narrow question: when a large language model offers an “explanation,” does it amount to a reason in Brandom’s sense, or is it only reason-shaped prose? In the small case examined here, the model’s answers have the surface form of reason-giving: they cite considerations, acknowledge risks, and respond intelligibly when pressed. The gap appears when incompatibility is introduced. Under vocabulary pressure, the model absorbs morally loaded predicates without letting them force a clear inferential shift. Under incompatibility pressure, it can describe what revision would require, but it does not cleanly own a withdrawal and then carry the downstream inferential costs forward.

The verdict is deliberately modest: in this case, the model’s explanations function as tools that can be taken up inside our space

9 Brandom, *Tales of the Mighty Dead*, 350–51, 360–61.

10 Charlotte Blease et al., “Generative Artificial Intelligence in Primary Care,” *BMJ Health & Care Informatics* 31, no. 1 (2024): e100872; Camille Grévisse, “LLM-Based Automatic Short Answer Grading in Undergraduate Medical Education,” *BMC Medical Education* 24, no. 1 (2024): 1060; *Humanities and Social Sciences Communications*, “The Effects of ChatGPT on Learning Performance” (2025).



of reasons, but they do not yet appear as unambiguous participants within it.

Future work could apply the same probes to other models, longer exchanges, and different kinds of explanation in order to determine whether similar patterns of revision and non-withdrawal persist.





Mohiz (he/him) is a third-year undergraduate at the University of Windsor in Canada, pursuing a double major in philosophy and information technology. His interests lie at the intersection of metaethics, the metaphysics of consciousness, and the philosophy of AI, alongside a lasting affection for poetry and for questions about the moral fate of minded beings, actual and possible.



THE CASE FOR AI AUTHORITY IN DIAGNOSTIC DECISIONS

ABSTRACT

This paper argues that artificial intelligence should transition from a supplementary tool to a central authority in medical diagnosis and treatment. While human physicians have traditionally held epistemic authority, their decision-making is constrained by cognitive limits and unequal access to medical knowledge. I examine the ethical implications of this shift through principles of beneficence and epistemic justice, arguing that clinical authority should follow the most reliable source of knowledge. In conclusion, I propose a hybrid model in which AI guides diagnosis, while physicians retain responsibility for ethical oversight and patient-centered care.

ROHIL WATWE



I. INTRODUCTION

Artificial intelligence (AI) has made significant strides in the medical field, specifically in disease diagnosis. As AI continues to improve, the role of human physicians as the centerpiece of the medical system has come into question. Traditional beliefs hold that human physicians are better equipped to be the final decision-makers due to their experience and ability to reason ethically. However, with AI demonstrating superior diagnostic capabilities in various medical fields, its role is shifting from a supplementary tool to a potential central authority in clinical decision-making.¹

I argue that, as AI systems evolve, they will be better equipped to make clinical decisions than human doctors, and thus, should transition from a supplementary tool to a central authority in diagnosis and treatment. This argument rests on three claims. Firstly, AI's superior diagnostic accuracy makes deference to less reliable human judgment ethically problematic. Secondly, AI has access to our entire collection of medical data, surpassing human limitations and allowing for more consistent and equitable care. Third, while both AI and humans are susceptible to bias, AI's decision-making processes are correctable in ways that human cognition is not, making it a more ethically accountable tool.

The increasing efficacy of AI in disease diagnosis highlights its potential to revolutionize healthcare. AI has the ability to process data on a large scale, identifying patterns that humans cannot. Furthermore, AI can refine its accuracy by integrating real-time patient data. In contrast, humans are limited in this scope, affected by cognitive constraints and memory limitations. AI's ability to synthesize medical information at a rapid rate gives it a distinct advantage over human practitioners.² Additionally, human physicians are susceptible to bias and fatigue, which can impact diagnostic accuracy, whereas properly designed AI applies standardized criteria to all cases, allowing for greater objectivity in clinical decision-making. While critics say that AI can inherit bias based on training data, efforts to improve training dataset diversity should mitigate these risks.³

1 Andre Esteva et al., "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks," *Nature* 542, no. 7639 (February 2017): 115–18, <https://doi.org/10.1038/nature21056>.

2 Chayakrit Krittanawong et al., "Artificial Intelligence in Precision Cardiovascular Medicine," *Journal of the American College of Cardiology* 69, no. 21 (May 2017): 2657–64, <https://doi.org/10.1016/j.jacc.2017.03.571>.

3 Ziad Obermeyer et al., "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations," *Science* 366, no. 6464 (October 2019): 447–53, <https://doi.org/10.1126/science.aax2342>.



Critics argue that AI lacks fundamental humanizing characteristics that make human physicians particularly effective. These include traits such as empathy, ethical reasoning, and patient communication. Especially in end-of-life care, humans provide sensitivity and judgment that AI can struggle to replicate. Although these limitations do exist, they do not diminish the power of AI as a diagnostic and decision-making tool. Rather than serving only a supplementary role, AI should move toward central authority in diagnosis and treatment, while human physicians maintain ethical oversight and patient-centered care.

II. HISTORICAL CONTEXT

The use of AI in a medical context is not a new phenomenon. In the 1970s, a tool known as MYCIN was used to identify bacteria causing severe infections.⁴ This system ran on the earliest iterations of AI. In the decades since, AI has improved substantially, both in diagnostic capability and efficiency. With the creation of machine learning algorithms, AI has become more capable of recognizing complex patterns in large datasets.

Modern AI systems use neural networks trained on millions of images and lab results. These systems can detect subtle indications of disease that humans are prone to miss. In fact, many modern specialties, including radiology, dermatology, pathology, and ophthalmology, utilize AI systems for pattern recognition, leading to enhanced patient outcomes.⁵

In addition to pattern recognition, AI also has the ability to outperform human physicians. A 2017 study by Ranav Rajpurkar et al. used a deep learning model to help detect pneumonia on chest X-rays.⁶ When comparing performance with human physicians, the AI was able to outperform humans in detecting fourteen different diseases. Similarly, in a study done through Google's DeepMind AI system, artificial intelligence was found to detect over fifty retinal diseases with an

4 Edward H. Shortliffe, *Computer-Based Medical Consultations: MYCIN* (Elsevier, 1976), <https://doi.org/10.1016/B978-0-444-00179-5.50001-9>.

5 Geert Litjens et al., "A Survey on Deep Learning in Medical Image Analysis," *Medical Image Analysis* 42 (December 2017): 60-88, <https://doi.org/10.1016/j.media.2017.07.005>.

6 6 Pranav Rajpurkar et al., "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning," *arXiv* (2017), <https://doi.org/10.48550/arXiv.1711.05225>.



accuracy comparable to that of leading ophthalmologists.⁷ There is an abundance of data that points to AI being able to outperform human physicians. These deep learning models also benefit from being trained on current medical data continuously, meaning diagnostic accuracy should keep improving over time.

III. LIMITATIONS OF HUMAN DECISION-MAKING

While physicians have extensive knowledge and clinical experience, their decision-making is prone to limitations. Cognitive biases have been shown to affect the present medical system deeply. Diagnostic errors affect an estimated 10–15% of medical cases, with cognitive factors contributing substantially to these mistakes.⁸ These errors reflect recurring vulnerabilities in human reasoning under uncertainty rather than isolated incidents. For example, confirmation bias can lead physicians to settle on an initial diagnosis prematurely and overlook contradictory evidence, increasing the risk of misdiagnosis.⁹ Additionally, physicians tend to work long hours, exacerbating issues such as physician fatigue. A 2004 study by Christopher P. Landrigan et al. found that diagnostic errors increased by around 20% during extended shifts in the emergency department.¹⁰

Physicians also face the challenge of staying up to date with current medical knowledge. One study asserts that in 2020, the doubling time of medical knowledge was around 73 days.¹¹ It is unrealistic for physicians to stay up to date on every advancement, something AI can accomplish with far fewer issues. With humans facing so many limitations, having AI as a secondary authority seems to be even more problematic.

7 Jeffrey De Fauw et al., “Clinically Applicable Deep Learning for Diagnosis and Referral in Retinal Disease,” *Nature Medicine* 24, no. 9 (September 2018): 1342–50, <https://doi.org/10.1038/s41591-018-0107-6>.

8 David E. Newman-Toker et al., “Burden of Serious Harms from Diagnostic Error in the USA,” *BMJ Quality & Safety* 33, no. 2 (January 2024): 109–20, <https://doi.org/10.1136/bmjqs-2021-014130>.

9 Pat Croskerry, “The Importance of Cognitive Errors in Diagnosis and Strategies to Minimize Them,” *Academic Medicine* 78, no. 8 (August 2003): 775–80, <https://doi.org/10.1097/00001888-200308000-00003>.

10 Christopher P. Landrigan et al., “Effect of Reducing Interns’ Work Hours on Serious Medical Errors in Intensive Care Units,” *New England Journal of Medicine* 351, no. 18 (October 2004): 1838–48, <https://doi.org/10.1056/NEJMoa041406>.

11 Peter Densen, “Challenges and Opportunities Facing Medical Education,” *Transactions of the American Clinical and Climatological Association* 122 (2011): 48–58.



IV. THE PROBLEM TO BE SOLVED

Despite evidence suggesting AI's growing capabilities in diagnostic tasks, human physicians remain the final decision-makers in diagnosis and treatment. This status quo is perpetuated due to beliefs that humans have professional expertise and provide ethical reasoning. However, by focusing solely on diagnostic tasks, where the ultimate goal is accuracy, the deference to human judgment is increasingly difficult to defend. Diagnostic error is one of the leading causes of preventable harm in healthcare. It is estimated that approximately 795,000 Americans suffer permanent disability or death each year due to diagnostic errors.¹² When cognitive limitations contribute to harm on this scale, the ethical implications of preserving exclusive human authority become significantly heightened.

This raises an important question: Should AI, which has been demonstrated to be superior at recognizing patterns and processing large quantities of medical data, remain behind the judgment of human physicians in diagnosis? While AI does not replace the need for empathy and ethical reasoning, it provides many other benefits. In places where accuracy and consistency are critical, such as when detecting diseases and interpreting scans, AI has proven to be more effective than humans. The issue now is not whether AI can match human diagnostic performance but rather when the medical system is ready to acknowledge the advantages of AI and reorganize accordingly.

V. DIAGNOSTIC ACCURACY AND ETHICAL AUTHORITY

To justify AI's transition from a supplementary tool to a central authority, it is not enough to demonstrate its superior diagnostic accuracy. One must also show that this superiority carries ethical weight. Specifically, in an epistemic and moral framework, clinical authority should follow the most reliable source of medical knowledge. If AI consistently outperforms human physicians in diagnostic accuracy, then continuing to emphasize less accurate human judgment becomes ethically indefensible. For instance, the landmark study by Andre Esteva et al. demonstrated that deep neural networks matched the performance of dermatologists when classifying skin cancer.¹³ A 2019 study by Maged Nasser and Umi Kalsom Yusof showed

¹² Newman-Toker et al., "Burden of Serious Harms," 109–20.

¹³ Esteva et al., "Dermatologist-Level Classification," 115–18.



that an AI system was the most accurate and extensively used model for breast cancer detection.¹⁴ These are not isolated successes but rather indicative of a broader trend in AI-driven diagnostics.

These implications go far beyond performance metrics. The concept of epistemic injustice asserts that patients are wronged when their access to knowledge, including life-saving medical information, is obstructed by institutional structure.¹⁵ Prioritizing the authority of human physicians, even when superior epistemic tools exist, is an ethical failure. It privileges the less accurate information of a human over a better-informed source, undermining a patient's right to the best knowledge available. Importantly, this argument concerns the ethical duty to deliver accurate clinical information, not the broader human responsibility of emotional support, which still rightly belongs to human physicians.

If the ethical foundation of medicine is, as Tom L. Beauchamp and James Childress argue, based on principles such as beneficence and nonmaleficence, then human physicians are under an obligation to provide the most accurate care available.¹⁶ With this in mind, clinical authority must follow epistemic authority. Choosing human judgment for diagnosis over more accurate options goes against this moral responsibility.

VI. ACCESS TO DATA AND EPISTEMIC SUPERIORITY

A core reason AI should have the final say in diagnosis lies in its unparalleled access to medical knowledge, giving it epistemic superiority over human physicians and providing consistent care. An individual physician's training is strongly variable. They each have different educations, different clinical experiences, and the aforementioned limitations in memory and cognition. Unlike humans, AI systems have been consistently designed to interpret large amounts of medical data with precision.

This capacity goes beyond data processing. AI functions as a sort of collective intelligence. For example, large language models like

14 Maged Nasser and Umi Kalsom Yusof, "Deep Learning Based Methods for Breast Cancer Diagnosis: A Systematic Review and Future Direction," *Diagnostics* 13, no. 1 (January 2023): 161, <https://doi.org/10.3390/diagnostics13010161>.

15 Miranda Fricker, *Epistemic Injustice: Power and the Ethics of Knowing* (Oxford University Press, 2007), <https://doi.org/10.1093/acprof:oso/9780198237907.001.0001>.

16 Tom L. Beauchamp and James F. Childress, *Principles of Biomedical Ethics*, 7th ed. (Oxford University Press, 2013).



DeepMind's AlphaFold or IBM Watson Health represent a collection of thousands of medical studies and clinical data, far beyond what a single human could replicate. This epistemic advantage has both practical and moral implications. If the goal of a medical system is to maximize health outcomes, then using AI offers higher accuracy, faster throughput, and more consistent care.

Crucially, AI's scalability allows a "leveling of the playing field." Human expertise is generally unevenly distributed. Big cities might have hundreds of physicians and many subspecialists, while rural areas tend to suffer from physician shortages. AI offers a standardized diagnostic tool accessible to all areas. Every patient, regardless of socioeconomic status or geographic location, can benefit from a "best possible" doctor whose expertise is not affected by human limits. AI's superior access to data is not just a technical strength. Rather, it forms an epistemic foundation for centralized diagnostic authority. In a system that emphasizes equity, it is wrong to limit medical care to individual physicians when we have tools that provide a more complete view of the healthcare field.

VII. BIAS AND MORAL RESPONSIBILITY

Bias is central to the debate over medical authority, and AI holds the ethical advantage. This is not because AI is bias-free but because its biases are detectable and thus more manageable than those of human physicians. Human physicians, despite their extensive training, are susceptible to implicit biases. These have been shown to produce systemic inequalities in healthcare. For instance, a study in 2007 by Alexander Green et al. found that physicians demonstrated unconscious pro-white biases that affected clinical decisions.¹⁷ Kelly M. Hoffman et al. documented that Black patients were less likely to receive pain medication due to false beliefs about biological differences.¹⁸

With regard to bias, a counterpoint that many critics mention is that AI systems have inherent biases that are directly comparable to those of their human counterparts due to biased training data. This critique conflates two distinct issues: the presence of bias and the cor-

17 Alexander R. Green et al., "Implicit Bias among Physicians and Its Prediction of Thrombolysis Decisions for Black and White Patients," *Journal of General Internal Medicine* 22, no. 9 (September 2007): 1231-38, <https://doi.org/10.1007/s11606-007-0258-5>.

18 Kelly M. Hoffman et al., "Racial Bias in Pain Assessment and Treatment Recommendations, and False Beliefs about Biological Differences between Blacks and Whites," *Proceedings of the National Academy of Sciences* 113, no. 16 (April 2016): 4296-4301, <https://doi.org/10.1073/pnas.1516047113>.



rectability of bias. While both humans and AI systems may exhibit bias, the moral and practical significance lies in whether these biases can be recognized and fixed.

The human brain is opaque. Implicit biases operate at the subconscious level, making them difficult to detect and even harder to change. Attempts at bias training and awareness programs have yielded mixed results and tend to wear off over time.¹⁹ In contrast, AI systems operate through code and models that are open to inspection. As Joy Buolamwini and Timnit Gebru demonstrated in their work on facial recognition AI bias, algorithmic disparities can be systematically detected and quantified.²⁰

This distinction has great ethical implications. A just healthcare system should strive to not only reduce morally arbitrary disparities in treatment but also correct existing inequalities. AI offers a path toward this ideal, not because it is inherently unbiased but because its mechanism of decision-making is transparent and subject to external regulation. Tools such as fairness-enhancing algorithms and reweighting algorithms allow for continuous adjustment and improvement. In contrast, efforts to reform human decision-making tend to rely on individual behavioral change, which faces significant inertia.

In short, the question is not about whether bias exists in AI. Rather, it is about whether we can do something about it. In this aspect, AI offers the possibility of continuous ethical progress that human physicians lack.

VIII. ADDRESSING COUNTERARGUMENTS

A common objection to AI having the final say in medical diagnosis stems from the idea that AI lacks the capacity for empathy and moral reasoning. Human physicians offer emotional support and comfort, which is particularly useful for end-of-life care. Critics say that AI cannot possibly hope to replicate the bond and trust that patients develop with their doctors.²¹ This concern is valid and important, par-

19 Temitayo A. Ogunleye, "Unconscious Bias," *Dermatologic Clinics* 41, no. 2 (April 2023): 285–90, <https://doi.org/10.1016/j.det.2022.08.003>.

20 Joy Buolamwini and Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," *Proceedings of Machine Learning Research* 81 (2018), <https://proceedings.mlr.press/v81/buolamwinil8a.html>.

21 Mary Catherine Beach et al., "Relationship-Centered Care: A Constructive Reframing," *Journal of General Internal Medicine* 21, no. S1 (January 2006): S3–8, <https://doi.org/10.1111/j.1525-1497.2006.00302.x>.



ticularly for patient-centered care where communication and compassion are deeply interrelated with treatment outcomes.²² Since AI lacks consciousness, critics state that it cannot replicate the nuanced decisions a human makes on a day-to-day basis. However, this thought misunderstands my argument. I do not suggest that AI should replace all human roles in medicine, but rather that AI should guide the technical core of medical decision-making, including diagnosis and treatment recommendations. AI's utility lies in improving clinical precision, not in replacing the human presence in care. A hybrid model is necessary. AI systems should guide diagnostic and treatment planning while physicians take responsibility for patient communication and ethical counseling.

Recent studies have shown that a hybrid model is already proving effective. A 2022 study by Khalid Al-Dasuqi et al. showed that AI systems that flag possible abnormalities are best used when their outputs are reviewed by human radiologists for accuracy and consistency.²³ The physician then interprets these findings within the context of the patient's complete medical history and clinical presentation. This division of labor strengthens the doctor-patient relationship by allowing the physician to focus on qualities only they can provide such as compassion and trust.

Another objection to AI systems is that they are susceptible to diagnostic errors. There have been instances of false positives, false negatives, and blind spots in AI training algorithms. In high-stakes areas of medicine, even a small margin of error could contribute to devastating consequences. This is a serious concern. AI is not infallible, and overconfidence in its abilities could lead to dangerous outcomes. However, this objection underestimates the detectability and the correctability of AI errors. AI systems leave a trail of logic and data behind their conclusions that can be iteratively improved through retraining, something that human cognition struggles to do. Moreover, comparative studies show that AI systems have the same rate of diagnostic error as human physicians in several fields. For example, a study in *Nature Medicine* showed that an AI model diagnosed pediatric illnesses correctly at a rate comparable to senior consultants.²⁴ The broader

22 Hui-Ching Weng et al., "The Effect of Surgeon Empathy and Emotional Intelligence on Patient Satisfaction," *Advances in Health Sciences Education* 16, no. 5 (December 2011): 591-600, <https://doi.org/10.1007/s10459-011-9278-3>.

23 Khalid Al-Dasuqi, Michele H. Johnson, and Joseph J. Cavallo, "Use of Artificial Intelligence in Emergency Radiology: An Overview of Current Applications, Challenges, and Opportunities," *Clinical Imaging* 89 (September 2022): 61-67, <https://doi.org/10.1016/j.clinimag.2022.05.010>.

24 Huiying Liang et al., "Evaluation and Accurate Diagnoses of Pediatric Diseases Using Artificial Intelligence," *Nature Medicine* 25, no. 3 (March 2019): 433-38, <https://doi.org/10.1038/s41591-018-0335-9>.



context matters here. Diagnostic error is one of the most common and costly types of medical error globally, with around 795,000 Americans becoming permanently disabled or dying annually because of it.²⁵ Even if AI systems are imperfect, they are systematically improvable. By leveraging large-scale data, AI is uniquely equipped to evolve in accuracy over time.

A final consideration lies in ethical accountability. If an AI system misdiagnoses a patient, who should take responsibility: the hospital, the AI developer, or the physician? This ambiguity is a big reason why many resist giving AI systems authority in the medical field. This concern is also justified. The legal infrastructure surrounding AI in medicine is currently underdeveloped. However, this is not a reason to exclude AI from decision-making. Rather, it should serve as a wake-up call to accelerate legal and ethical reform in the field.

Already, people are proposing frameworks for shared liability between developers and physicians.²⁶ These proposals generally follow a distributed responsibility model. In such a model, AI developers are responsible for the design, training data integrity, validation procedures, and post-market monitoring of their systems. Hospitals and healthcare institutions are responsible for proper implementation, oversight, and ensuring that AI tools are used within their validated scope. Physicians remain responsible for integrating AI outputs into the broader clinical picture and communicating decisions to patients. Liability is therefore not placed on a single actor but allocated according to control, expertise, and the capacity to prevent harm.

Within the hybrid model I propose, responsibility would follow function. AI systems hold epistemic authority in diagnosis as far as they provide the most accurate probabilistic assessment of disease. However, physicians would retain ultimate relational and ethical responsibility. This means that while AI may generate diagnostic conclusions, physicians remain accountable for final clinical implementation, contextual interpretation, and informed consent discussions. If harm arises from flawed model design or biased training data, responsibility would extend up to developers and regulatory bodies. If harm arises from improper use or misinterpretation of AI output, responsibility would lie with the clinical institution or physician.

This functional allocation reflects the operational structure of the hybrid model itself. Authority is divided according to comparative

25 Newman-Toker et al., "Burden of Serious Harms," 109–20.

26 Sara Gerke, Timo Minssen, and I. Glenn Cohen, "Ethical and Legal Challenges of Artificial Intelligence-Driven Healthcare," in *Artificial Intelligence in Healthcare*, ed. Adam Bohr and Kaveh Memarzadeh (Elsevier, 2020), 295–336, <https://doi.org/10.1016/B978-0-12-818438-7.00012-5>.



strength: AI maximizes diagnostic reliability, while human physicians maintain ethical stewardship and patient-centered care. Accountability must mirror that same division. The presence of distributed responsibility does not weaken ethical clarity; rather, it acknowledges the collaborative nature of modern medical systems.

Ultimately, the potential benefits of AI-guided diagnosis outweigh the legal uncertainties. The challenges mentioned above are solvable, especially when compared against the status quo, where human error remains the third leading cause of death in the US.²⁷

IX. CONCLUSION

The integration of artificial intelligence into clinical diagnosis should be considered a moral necessity. This paper has argued that, given AI's superior diagnostic ability, unparalleled access to comprehensive medical data, and capacity for bias correction, it should be given the final say in the diagnostic process. Deferring to human judgment for diagnosis represents a failure to uphold key principles of beneficence and nonmaleficence.

Importantly, this claim does not require replacing human physicians. Humans provide qualities such as empathy and ethical needs that are essential features of medical care, something an AI system cannot and should not replicate. Rather, I argue for a hybrid model in which AI provides the epistemic foundation for diagnosis and treatment, and human physicians focus on patient relations and the ethical dimension of care. This division of labor will allow both parties to play to their strengths.

Additionally, the expectation that AI will continue to improve diagnostically is not speculative optimism, but an inference supported by current trends in machine learning. AI systems benefit from increased volumes of high-quality health data and continual advancements in model architecture. For instance, recent developments allow AI models to be trained across multiple hospitals without compromising patient privacy, accelerating real-world improvement.²⁸ Moreover, regulatory bodies like the FDA have already approved adaptive AI systems that update themselves post-deployment, reinforcing the idea

27 Martin A. Makary and Michael Daniel, "Medical Error—The Third Leading Cause of Death in the US," *BMJ* (May 2016): i2139, <https://doi.org/10.1136/bmj.i2139>.

28 Nicola Rieke et al., "The Future of Digital Health with Federated Learning," *NPJ Digital Medicine* 3 (2020): 119, <https://doi.org/10.1038/s41746-020-00323-1>



that diagnostic accuracy by AI is not static but subject to continual refinement.

Objections about errors in AI diagnosis and the unclear nature of legal accountability are valid concerns. However, they are not disqualifying. Diagnostic errors already cost thousands of lives annually under human authority. Unlike human error, AI errors are traceable and improvable. Additionally, developing frameworks for shared legal responsibility is underway.

In conclusion, the status quo of human physicians having the final decision in the medical process cannot be ethically justified when more reliable systems are available. If medicine is to evolve with its ethical commitments, it must embrace AI not just as a tool but as a diagnostic authority. To delay this shift is to allow outdated norms to obstruct life-saving progress.







Rohil Watwe is a junior at Duke University in Durham, North Carolina, studying biology with a minor in chemistry and a certificate in innovation and entrepreneurship. His philosophical interests include bioethics and medical decision-making, with an emphasis on the role of artificial intelligence in healthcare



SINGER'S SHALLOW POND THOUGHT EXPERIMENT AND VIRTUE ETHICS

ABSTRACT

Peter Singer's shallow pond thought experiment is often used to argue that we have a moral obligation to provide aid to others, regardless of factors such as proximity and number of bystanders. Critics claim that Singer's thought experiment fails to capture key features of global poverty and should thus not be applied to analyses of global justice. I argue that this use of the shallow pond thought experiment can be maintained only if we reconsider the thought experiment as entailing a virtue ethics principle of moral obligation. I support this conclusion by introducing a modified 'never-ending' shallow pond thought experiment.

JACK RAGAN



I. INTRODUCTION

Peter Singer's 1972 article *Famine, Affluence and Morality* is seen as being of significant importance in the field of global justice, contributing to discussions on poverty, global aid, and morality.¹ In the article, Singer develops following the shallow pond thought experiment:

On your way to work, you pass a small pond. On hot days, children sometimes play in the pond, which is only about knee deep. The weather's cool today, though, and the hour is early, so you are surprised to see a child splashing about in the pond. As you get closer, you see that it is a very young child, just a toddler, who is flailing about, unable to stay upright or walk out of the pond. You look for the parents or babysitter, but there is no one else around. The child is unable to keep his head above the water for more than a few seconds at a time. If you don't wade in and pull him out, he seems likely to drown. Wading in is easy and safe, but you will ruin the new shoes you bought only a few days ago and get your suit wet and muddy.²

Singer uses the shallow pond thought experiment to argue that in certain cases we have a moral obligation to act and provide aid to others, and this moral action should not be confined by factors such as proximity and number of bystanders.³ The shallow pond thought experiment is generally interpreted by both critics and proponents of Singer's work as an argument for a utilitarian solution to global poverty and issues of global justice.⁴

In this paper, I argue that Singer's shallow pond thought experiment is best interpreted as an argument for a virtue ethics solution to global poverty and issues of global justice. To support my thesis, I will first provide a modified version of Singer's shallow pond thought experiment, with the modifications providing reasons as to why we should interpret the thought experiment as arguing for virtue ethics. Secondly, I will discuss the theoretical implications of interpreting the shallow pond thought experiment as an argument for virtue ethics, as well as how this interpretation better reflects the cyclical nature of poverty. I will address how it provides a more practical solution to the bystander effect which commonly arises in global aid and crisis

1 Peter Singer, "Famine, Affluence and Morality," *Philosophy and Public Affairs* 1, no. 3 (1972): 231-233, <https://www.jstor.org/stable/2265052>.

2 Peter Singer, *The Life You Can Save* (Random House, 2009), 3.

3 Singer, "Famine, Affluence and Morality," 231-33.

4 Andrew Kuper, "More Than Charity: Cosmopolitan Alternatives to the Singer Solution," *Ethics and International Affairs* 16, no. 1 (2002): 109.



resolution efforts.⁵ Accordingly, I argue that the principle developed through my modified shallow pond thought experiment, if implemented, would result in greater helping behaviors compared to Singer's original principle, since my principle better reflects the global material conditions (e.g., severity and nature of global injustices) and relevant social relations and phenomena (e.g., the bystander effect) that fail to be captured by Singer's shallow pond thought experiment.⁶ To support this line of reasoning I will draw upon psychological and philosophical research in the field of applied ethics.

II. SINGER'S SHALLOW POND THOUGHT EXPERIMENT

Before presenting a modified shallow pond thought experiment it is important to better understand the common interpretations of the original shallow pond thought experiment. Singer interprets the shallow pond thought experiment as presenting a utilitarian argument or principle: that we have a moral obligation to help others, or prevent something bad from happening, so long as we are not sacrificing something of comparable moral importance.⁷ In the context of global poverty, this means that the drowning child represents the global poor and the shallow pond represents the negative and potentially deadly impacts of poverty. Singer argues that we are morally obligated to save the life of the child or provide aid to the poor so long as we are not sacrificing something of comparable moral importance. In the case of the thought experiment, this means dirtying our clothes and, in the case of global poverty, donating a small amount of money to provide famine relief.⁸ While Singer's article focuses on other mod-

5 Johnathan Wolff, "Poverty," *Philosophy Compass* 14, no. 12 (2019): 4–5; Jon Mills, "The Global Bystander Effect: Moral Responsibility in Our Age of Ecological Crisis," *Journal of Future Studies* 25, no. 2 (2020): 67–68; Eliezer Yudkowsky, "Cognitive Biases Potentially Affecting Judgement of Global Risks," in *Global Catastrophic Risks*, ed. Nick Bostrom and Milan M. Ćirković (Oxford University Press, 2011), 110–12.

6 Emiliano Barrera et al., "Examining COVID-19 mortality inequalities across 169 countries: Insights from the COVID-19 mortality inequality curve (CMIC) and Theil index analysis," *Healthy Policy* 157 (2025): 5; United Nations Development Programme, "2021/22 Human Development Report," United Nations Development Programme, accessed December 12, 2023, <https://report.hdr.undp.org/intro>. Although global wealth inequality is decreasing, events like the COVID-19 pandemic demonstrate that this perceived increase in equality of wealth and income is neither fixed, nor necessarily correlated with well-being. Furthermore, although income inequality between equality is improving, the income equality within countries is worsening.

7 Singer, "Famine, Affluence and Morality," 231.

8 Singer, "Famine, Affluence and Morality," 235.



erators to moral obligations like proximity and number of bystanders, the shallow pond, experiment, as interpreted by Singer, primarily provides an argument for a utilitarian principle which serves to guide ethical decision making.⁹

Many of Singer's critics agree with the philosophical implications of the thought experiment, with much of the criticism of the shallow pond thought experiment being politically oriented. Critics argue that shallow pond thought experiment fails to provide a complete picture of the systemic and structural causes of global poverty, arguing that the shallow pond thought experiment oversimplifies global poverty and creates a white or western savior narrative.¹⁰ While there are a number of ways to understand the distinction between western savior and grassroots movements, in this paper I will employ the following two-fold distinction: Firstly, there is an epistemic difference between western savior movements and grassroots movements. Western savior movements employ a 'top-down,' expert-driven approach, which seeks to reinforce, either directly or indirectly, dominant and oppressive structures and narratives while failing to take into account the diverse epistemologies and standpoints of the very groups and individuals they are attempting to aid.¹¹ Conversely, grassroots movements do seek to overcome these dominant and oppressive structures and narratives by involving the local communities and groups which they are attempting to aid through a 'bottom-up' approach, thus ensuring that diverse ways of knowing and unique lived experiences are given 'deference' or epistemic and normative authority.¹² Secondly, and arising from this previous distinction, western savior movements differ from grassroots movements in their practical effectiveness. Western savior movements fail to interrogate or dismantle oppressive structures and systems; thereby, they are unable to create long-term change. Conversely, grassroots movements are able to create long-term change through their involving the unique standpoints of local communities, groups

9 Singer, "Famine, Affluence and Morality," 240–41.

10 Scott Wisor, "Against shallow ponds: an argument against Singer's approach to global poverty," *Journal of Global Ethics* 7, no. 1 (2011): 19, <https://doi.org/10.1080/17449626.2010.548819>.

11 Boaventura de Sousa Santos, *Epistemologies of the South* (Routledge: 2014), viii–x.

12 Ann Hill, Justin See and The Grassroots Collective, "Grassroots modalities of learning: creating and maintaining pluriversal knowing, being and doing for collective survival," *The International Journal of Justice and Sustainability* 30, no. 7 (2025): 831–34; Laurence Thomas, "Moral Deference," in *Theorizing Multiculturalism: A Guide to the Current Debate*, ed. Cynthia Willett (Blackwell, 1998), 372–78.



and individuals, as evidenced by the success of a number of grassroots movements.¹³

If we are to interpret the shallow pond thought experiment exactly as Singer has, these issues become evident. In the thought experiment, the western individual, who is positioned to provide aid to the global poor, is depicted as the sole agent capable of providing aid to the drowning child. The global poor are depicted as small children, unable to lift themselves out of poverty.¹⁴ In response to these criticisms, philosophers like Monique Deveaux provide competing models for addressing poverty, with Deveaux valuing grassroots movements and input from the global poor in order to shape solutions to global poverty.¹⁵ Singer's shallow pond, as well as much of *Famine, Affluence, and Morality*, ignores these criticisms of Singer's approach to poverty and solutions to global poverty as being grounded in a western savior narrative and a 'top-down' approach, rather than a grassroots 'bottom-up' approach.

However, it is important to contextualize the shallow pond thought experiment as being focused on developing an ethical principle, which can then be applied to political issues. While there are many grounds for political objections to Singer's shallow pond, it is important to remember that ethical claims, specifically that of having an obligation to prevent harm if we can do so by sacrificing something of lesser moral importance, are what grounds Singer's political thinking in *Famine, Affluence and Morality*.¹⁶ Therefore, if we object to the political implications of the shallow pond argument, we also object to some of the ethical claims which ground the argument. In order to refute the political implications of Singer's shallow pond we must provide objections to the original and grounding ethical claim. When critics object to the depictions of poverty and the savior narrative present in *Famine, Affluence and Morality*, they are, at a fundamental level, objecting to the utilitarian philosophy of Singer, which is not concerned with these wholistic accounts of poverty, aiming solely to provide guiding principles which describe our moral duties to maximize the good generated by our actions.¹⁷ The inability of utilitarianism to provide complete accounts of moral decision making and

13 Herdi Sahrasad, Al Chaidar, and Maksum Syam, "A reflection on a peripheral movement; The "Save Aru" social movement 2013-2015 from a historical perspective," *Wacana* 20, no. 3 (2019): 560-82.

14 Wisor, "Against shallow ponds," 21.

15 Monique Deveaux, "Philosophical Misframings of Poverty," in *Poverty, Solidarity, and Poor-Led Social Movements* (Oxford University Press, 2021), 56-57.

16 Singer, "Famine, Affluence and Morality," 239.

17 Deveaux, "Philosophical Misframings of Poverty," 56.



its political context is present in the shallow pond argument.¹⁸ In the following section, I argue that the shallow pond thought experiment should be interpreted as providing a virtue ethics principle to guide moral decision making, which better reflects the political and cultural agency of the global poor.

III. THE NEVER-ENDING SHALLOW POND THOUGHT EXPERIMENT

The primary philosophical objection to Singer's shallow pond is that it presents a moral duty to aid in a situation which consists of a singular interaction or problem, with only one child drowning who must be saved.¹⁹ Many individuals would be quick to save this child, or in the case of global aid, donate five dollars to the global poor, if it meant no further action was required. However, when discussing global poverty, we must remember that there are hundreds of millions of people who require aid; therefore, it is useful to revise the shallow pond thought experiment with the modified never-ending shallow pond thought experiment:

You save the child in the original shallow pond thought experiment and continue on your walk to work. Several minutes later, you pass a second pond in which another child is drowning. Unfortunately, there are several hundred million ponds on your walk to work, all with children drowning who require your assistance. Furthermore, once you reach the end of the hundreds of millions of ponds with drowning children, there are several billion ponds filled with children, who are not drowning, yet are unable to climb out of the pond and are stuck, in an uncomfortable, and miserable state.

This never-ending shallow pond thought experiment provides a more complete and accurate account of issues of global justice. It demonstrates the complexity of global aid and poverty, reflecting the reality of the several hundreds of millions of people suffering from immediate danger due to famine, disease and war.²⁰ Furthermore, as the thought experiment suggests, even if we save all of these individuals “drowning in ponds” through global aid efforts, there are still billions suffering

18 Walter Sinnott-Armstrong, “Consequentialism,” *The Stanford Encyclopedia of Philosophy*, October 4, 2023, <https://plato.stanford.edu/entries/consequentialism/>.

19 Travis Timmerman, “Sometimes there is nothing wrong with letting a child drown,” *Analysis* 75, no. 2 (2015): 208, <https://www.jstor.org/stable/24671242>.

20 United Nations, “Human Development Report.”



from complex problems, reducing their quality of life.²¹

Intuitively, we cannot expect one individual to spend their entire life saving children from ponds. While it seems reasonable to save a few children and continue on our way, Singer's principle suggests that we should continue saving children until we die. The only thing which holds greater moral significance than a life is another life. Once we consider that we could sacrifice our time by spending it saving children from ponds, it becomes difficult to assert that we are giving up something of comparable moral significance compared to the many thousands of lives that could be saved.²² In the case of the never-ending shallow pond thought experiment, Singer's stringent utilitarian principle can be reduced to an absurd demand.

Ultimately, we want our moral principles to, in some way, reflect our moral intuitions.²³ That is not to say we should ground moral arguments on intuition. Rather, the demands of Singer's original principle violate our moral intuitions in such a way that it reduces the principal's utility and legitimacy. If Singer's argument aims to demonstrate our moral obligations to provide aid to the global poor, there are other approaches that satisfy this demand, while providing a principle that is less stringent and takes into account the previously discussed political objections to the original shallow pond thought experiment.

The never-ending thought experiments provide evidence for a virtue ethics principle, describing our moral obligations in regard to global aid. This principle can be formulated as such: We have a moral obligation to help others so long as it is in a virtuous manner. Virtue ethics often discusses virtue as a matter of degree, with virtuous behavior or action being the mean between excessively "good" and "bad" behavior and action.²⁴ In the case of the never-ending shallow pond thought experiment, it would be morally impermissible to pass by all the drowning children leaving them to die. However, it seems that one would be displaying an excess of virtue by spending their entire lives dedicated to saving the drowning children, when they could instead address the root causes of the problem, by teaching children to swim or providing barriers around the ponds. These solutions would address the problem of drowning children in a more meaningful way and, if implemented correctly, would likely save more children than

21 United Nations, "Human Development Report."

22 Timmerman, "Sometimes there is nothing," 211.

23 Phillip Stratton-Lake, "Intuitionism in Ethics," *The Stanford Encyclopedia of Philosophy*, May 15, 2020, <https://plato.stanford.edu/entries/intuitionism-ethics/>.

24 Aristotle, *Nicomachean Ethics*, trans. Terence Irwin (Indianapolis: Hackett Publishing, 1985), 1107a1.



if one were to simply continue lifting them out of ponds. Ultimately, this approach would allow the virtuous individuals to, at a certain point, stop directly helping the never-ending succession of drowning children but still demand that we have a moral obligation to help in some way aside from immediate aid.

In the context of the never-ending shallow pond thought experiment, the virtue ethics principle does not seem to be as demanding as Singer's principle, aside from some amount of helping behavior. I would like to briefly acknowledge a possible objection that arises from the flexibility of the virtue ethics principle. Proponents of the utilitarian principle, including Singer, may argue that by not specifying when exactly we are no longer required to provide aid to the drowning children, the virtue ethics principle is too vague to be of normative use. There are two senses in which the proponent of the utilitarian principle might claim that the virtue ethics principle is vague. The first is what I call theoretical vagueness, or that the virtue ethics principle is formulated in such a way that does not distinguish it from other theoretical positions. Consider how a 'goodness' principle demonstrates this potential shortcoming: We have a moral obligation to do good. In response to this principle, the utilitarian might claim that the principle is in fact compatible with the utilitarian principle and, since it is too *practically* vague, we should reject it in favour of the utilitarian principle. Here, by practical vagueness, I mean roughly that the principle does not provide a sufficiently complex standard or guide for action, as should be illustrated by the 'goodness' principle which is a paradigmatic case of a practically vague principle.

Accordingly, in response to this objection the proponent of the virtue ethics principle must demonstrate that their principle is importantly distinct from the utilitarian principle and thereby not theoretically vague, in addition to providing a sufficiently complex standard or guide for action to avoid the charge of practical vagueness. Ultimately, after considering the theoretical and practical applications of the virtue ethics principle, in the context of global aid and global poverty, it seems that although the virtue ethics principle may be practically vague at times, it is still more practical than Singer's utilitarian principle. Below I discuss the theoretical and practical implications to provide further justification for my virtue ethics principal interpretation of the shallow pond thought experiment.



IV. VIRTUE ETHICS AND PROVIDING AID TO THE GLOBAL POOR

If we are to interpret the shallow pond thought experiment as supporting a virtue ethics principle, we must explore how this principle applies to poverty and other issues of global justice. In order to virtuously help the global poor, we must apply careful thinking to understand what the most effective course of action is in order to aid those suffering from the negative effects of poverty. There are similarities between Singer's original principle and this virtue ethics principle, namely that we attempt to do the most possible good through our actions. However, the virtue ethics principle does not imply that self-sacrifice is needed to achieve this good. Perhaps the virtuous individual will reflect and realize that by donating some insignificant amount of money they could save several lives; therefore, the virtue ethics principle would suggest that this is morally obligated. However, if one is so poor that they are unable to monetarily contribute to global aid efforts, perhaps this virtuous individual could participate in global aid efforts in other ways. The flexibility of this principle appears to be its strength and a response to the political objections that Singer's critics raise.²⁵

For instance, consider the case of an individual living in extreme poverty, Singer's principle does not seem to assign this individual a role in providing solutions to global poverty.²⁶ Generally speaking, this individual would have very little of moral significance that they could sacrifice in order to provide aid to others, or even themselves. Singer's principle does not appear to recognize that the global poor have agency and a role to play in developing and implementing solutions to global poverty.²⁷ This is not to say the global poor are responsible for the harm they are enduring, but rather they should have real and significant agency over potential solutions to poverty.²⁸ By obligating that individuals must act in a virtuous manner, the virtue ethics principle suggests that we should take into account the perspectives and voices of those we are aiding, and not impose our values or solutions onto the global poor. Virtuous solutions imply a process of collaboration when it comes to addressing poverty. To superimpose systems which do not reflect the cultural values or the agency of those we are attempting to aid is arrogant and prejudiced. It is not virtuous. Therefore, the virtue ethics principle not only addresses the issues raised by the never-end-

25 Wisor, "Against shallow ponds," 27–28.

26 Wisor, "Against shallow ponds," 21.

27 Deveaux, "Philosophical Misframings of Poverty," 56.

28 Deveaux, "Philosophical Misframings of Poverty," 56.



ing shallow pond, but also the core political objections to Singer's shallow pond.

Furthermore, the virtue ethics principle suggests that helping behaviors and principles which guide helping behaviors, can exist independent of morality. While morality and virtue are concepts that share similarities, virtue seems to transcend morality in ways that may be practically useful. Consider, for instance, the distinction that Bernard Williams employs between ethics, as being the broad domain which focuses on the study of the question of how one should live a good life, and morality which is the specific narrow modern subdiscipline of ethics which focuses on questions of moral responsibility, and universal or impartial rules.²⁹ As Williams asserts, those philosophers who focus too closely on morality often seem to lose any real interest in the perspectives of the human beings who must actually live with a moral problem.³⁰ In this sense, the virtue ethics principle may not be a moral principle, at least following the distinction provided by Williams. Instead, the virtue ethics principle may be an ethical principle which helps to guide us towards living a good life, while ensuring that others are able to lead these kinds of good lives.

Accordingly, this emphasis that the virtue ethics principal places on broader human flourishing may allow it to transcend morality in ways that are personally rewarding and practically advantageous. To illustrate, consider how, similar to deontological ethical theories, virtue ethics is primarily concerned with the actions of the agent, or the character of the agent, rather than the consequences of their actions.³¹ Moral psychologists have found that increased moral reasoning often results in an increased capacity for helping behaviors.³² That is, when we learn about morality, and consider what is moral and immoral we gain an increased capacity for and increased likelihood to engage in helping behaviors.³³ Virtue ethics seems to promote this capacity in greater ways than utilitarianism, especially in the context of global aid. Virtue ethics encourages individuals to look inwards and consider which actions are necessary in order to develop a virtuous character and be a "good" person. By contrast, utilitarianism is often concerned with external results and consequences.

29 Bernard Williams, *Ethics and the Limits of Philosophy* (Fontana Press, 1985), 4–8.

30 Donald McDonald, "The Use of Philosophy," *The Center Magazine*, November/December 1983, 40–49.

31 Aristotle, *Nicomachean Ethics*, 1106b16.

32 Cristina Villegas De Posada and Elvia Vargas-Trujillo, "Moral Reasoning and Personal Behavior: A Meta-Analytical Review," *Review of General Psychology* 19, no 4 (2015): 416–17, <https://doi.org/10.1037/gpr0000053>.

33 De Posada and Vargas-Trujillo, "Moral Reasoning and Personal Behavior: A Meta-Analytical Review," 418.



In the case of global poverty, it can be easy to ignore these negative consequences, especially for individuals who are not directly exposed to the reality of extreme poverty.³⁴ However, the internalized nature of virtue ethics forces individuals to consider that a lack of action not only presents negative consequences but also reflects an unvirtuous character. Individuals will be more likely to internalize moral action and reflect on which course of action is moral and correct if we teach and apply a virtue ethics principle. By forcing individuals to internally reflect on morality, we are forcing them to examine and consider their own behavior on a more frequent basis, which as studies have demonstrated, should lead to increased helping behavior.³⁵

V. CONCLUSION

Upon considering the theoretical and practical reasons for interpreting the shallow pond thought experiment as an argument for a virtue ethics principle of moral obligation, this approach provides a compelling middle ground for both Singer and his critics to consider. The virtue ethics principle provides a more complete and accurate account of global poverty and our moral obligations to act, suggesting that we have a moral obligation to help others so long as it is in a virtuous manner. The original shallow pond has now outlived its utility, and in an era where issues of global justice are becoming increasingly interconnected and frequent, it is evident that we must develop new approaches which reflect the complex nature of issues of global justice.³⁶ A virtue theory of ethics provides compelling answers to these problems and as demonstrated by the never ending shallow pond thought experiment, allows us to not only describe our obligations to the global poor but also to create solutions to issues of global justice.

34 David J. Franz, "Is Applied Ethics Morally Problematic," *Journal of Academic Ethics* 20, (2022): 359-74, <https://doi.org/10.1007/s10805-021-09417-1>.

35 De Posada and Vargas-Trujillo, "Moral Reasoning and Personal Behavior," 416-17.

36 Jonathan Wolff, "Method in Philosophy and Public Policy," in *The Routledge Handbook of Ethics and Public Policy*, ed. Annabelle Lever and Andrei Poama (Routledge, 2018), 17-23. See for example, the development of "Engaged Theorizing" or "Engaged Philosophy" movements, whose principles reflect those of the virtue ethics principle I propose.





Jack Ragan is a fourth-year philosophy student at Carleton University. His primary philosophical interests lie at the intersection of metaphysics and ethics. Jack is set to pursue a master's in philosophy at Carleton University, with his research focused on applying the capabilities approach to the ethics of AI.



CAN PHILOSOPHY HAVE A BEGINNING?

ABSTRACT

If philosophy began, how clear or obvious should we expect its origins to be? The historical argument contends that philosophy began with the shift to naturalistic explanation by the Presocratics. If philosophy did begin with them, we should expect to inhabit a different epistemic situation than we in fact do - one where we have a static understanding of the term. The evidence available for its start should be so definitive that it would be prudent for us to rule out what is and isn't philosophy, irrespective of the era. In this paper, I propose, through an examination of the problem, definition, Aristotle's work, and historical developments, a genealogical understanding of the origins of philosophy.

AARUSH PRABHU



I. INTRODUCTION

Can philosophy have a beginning? The question seems historical at first: it appears to ask for a date, a place, a first instance; however, it is not, because answering it requires us to determine what philosophy is. To ask when philosophy began is to ask whether philosophy can be defined and on what terms. If philosophy has a fixed criterion, we can search for it in its first instance; however, if it is an evolving practice, searching for a single origin risks mistaking retrospective classification for historical fact.¹

The dominating location of the beginning of philosophy is the Presocratics, whereby the shift from appealing to anthropomorphic gods to appealing to general, impersonal, and debatable naturalistic explanations marks the beginning.² Aristotle, among other authors, arranges earlier thinkers into a sequence of investigations of causes and principles, affirming the view that the Presocratics inaugurated something recognizably philosophical.³ The term “Presocratics” is a scholarly construction that misleads if taken as purely chronological since it suggests that these thinkers are merely preparatory, significant only for anticipating later philosophy.⁴ Furthermore, the term “philosophy,” derived from *philosophia*, did not designate a sharply defined discipline, but referred more broadly to a mode of intellectual engagement oriented toward inquiry and cultivation.⁵

This paper proposes a genealogical view whereby philosophy emerges when later thinkers recognize earlier inquiries as belonging to a specific field, thereby rendering the views of chronology moot to

-
- 1 Bernard Williams, “Philosophy as a Humanistic Discipline,” in *Philosophy as a Humanistic Discipline* (Princeton University Press, 2006); Alasdair MacIntyre, “The Relationship of Philosophy to Its Past,” in *Philosophy in History*, ed. Richard Rorty, J.B. Schneewind, and Quentin Skinner (Cambridge University Press, 1984).
 - 2 John Burnet, *Early Greek Philosophy* (A. & C. Black, 1892); F.M. Cornford, *From Religion to Philosophy* (Edward Arnold, 1912); G.E.R. Lloyd, *Magic, Reason and Experience* (Cambridge University Press, 1979).
 - 3 Aristotle, *Metaphysics*; Harold Cherniss, *Aristotle’s Criticism of Presocratic Philosophy* (Johns Hopkins Press, 1935); Jonathan Barnes, *The Presocratic Philosophers*, 2nd ed. (Routledge, 1982).
 - 4 Laks, *Concept of Presocratic Philosophy*; James Warren, *Presocratics* (University of California Press, 2007).
 - 5 Pierre Hadot, *Philosophy as a Way of Life*, trans. Michael Chase (Blackwell, 1995); Alexander Nehamas, *The Art of Living* (University of California Press, 1998). For early uses, see Herodotus, *Histories* 1.30; Thucydides, *History of the Peloponnesian War* 2.40.



acts of interpretation.⁶ This claim does not deny the importance of the Presocratics or the significance of the shift toward naturalistic explanation: it emphasizes that chronology alone cannot determine philosophical membership. What matters instead are the shifting standards through which explanations become intelligible, persuasive, and acceptable. It is these standards that allow later authors to distinguish myth from theology, science from philosophy, and to place earlier thinkers within or outside these categories. Before examining how such standards operate, we must first ask: what counts as philosophy?

II. THE PROBLEM OF DEFINITION

“To be or not to be?” is the question of definitions. Any claim about philosophy inherently presupposes an answer to what philosophy is: it distinguishes what it is from what it is not. To define philosophy, we must know which activities belong inside the category and which don’t, yet philosophy resists this separation. Its boundaries are difficult to specify. One might ask whether any reflective inquiry falls entirely outside philosophy or whether it emerges gradually from broader activity with no clear point of separation.

Using contemporary definitions to classify earlier thinkers’ definitions risks anachronism and presentism by projecting categories from modern institutions onto periods where they did not exist.⁷ Early Greeks did not operate within a discipline called philosophy. Applying the modern category to them presupposes that philosophy has a stable definition capable of extending across historical change.

For such an approach to succeed, philosophy would need an absolute definition that is fixed across time, yet many things change without losing their identity. A person remains the same individual despite constant physical and psychological change. Properties shift, but identity persists. Definitions can tolerate change without collapsing. The problem is whether philosophy has a comparable stability. Compare this with natural kinds. Consider water: earlier speakers described it as a clear, drinkable liquid, while modern chemistry identifies it as H₂O. The description changed, but the substance did not. The object remains stable as our understanding improves. Definitional re-

6 Friedrich Nietzsche, *On the Genealogy of Morality*, trans. Carol Diethe (Cambridge University Press, 1994); Michel Foucault, “Nietzsche, Genealogy, History,” trans. Donald F. Bouchard, in *Language, Counter-Memory, Practice* (Cornell University Press, 1977); Raymond Geuss, “Nietzsche and Genealogy,” *European Journal of Philosophy* 2, no. 3 (1994).

7 Quentin Skinner, “Meaning and Understanding in the History of Ideas,” *History and Theory* 8, no. 1 (1969).



finement does not imply instability in the object described.⁸ However, philosophy is different. Water exists independently of human description. Philosophy exists only through human activity. It is not a natural substance but a historically evolving practice. As traditions develop, meanings shift, boundaries change, and activities move between fields. Natural philosophy became physics. The study of the soul became psychology.⁹ What once belonged to philosophy can migrate into new disciplines as methods and institutions change.¹⁰ This creates a dilemma to define philosophy narrowly by modernity, and early thinkers are excluded because they lived before those standards developed. While defining philosophy broadly as reflective inquiry about fundamental questions would expand to include mythic cosmology, theological speculation, and poetic reflection, each definition reshapes what counts as the beginning of philosophy.

This argument recognizes the possibility that philosophy might have an underlying unity, but it is possible that its essence is not fully articulated. Irrespective of this, such an essence cannot be assumed in advance when determining the origins of philosophy.¹¹ Instead of searching for philosophy's first instance under a fixed definition, we can ask when earlier inquiries become intelligible as philosophy within later traditions. This shifts the focus from identifying a starting point to understanding how intellectual lineages are constructed. To see how such retrospective organization operates, we must examine the thinker who most shaped the familiar origin story: Aristotle.

III. RETROSPECTIVE LINEAGE AND THE MYTH-NATURE SHIFT

The case for locating philosophy's beginning with the Presocratics is rooted in Aristotle's way of writing intellectual history. In *Metaphysics*, Aristotle presents earlier thinkers as participants in a common inquiry into "causes and principles," Thales, Anaximander, and Anaximenes

- 8 Hilary Putnam, "The Meaning of 'Meaning'," in *Mind, Language and Reality: Philosophical Papers*, vol. 2 (Cambridge University Press, 1975); Saul Kripke, *Naming and Necessity* (Harvard University Press, 1980).
- 9 Andrew Cunningham, "Getting the Game Right: Some Plain Words on the Identity and Invention of Science," *Studies in History and Philosophy of Science* 19, no. 3 (1988); Gary Hatfield, "Remaking the Science of Mind: Psychology as Natural Science," in *Inventing Human Science*, ed. Christopher Fox, Roy Porter, and Robert Wokler (University of California Press, 1995).
- 10 Stephen Toulmin, *Human Understanding: The Collective Use and Evolution of Concepts* (Princeton University Press, 1972).
- 11 Jonathan Barnes, "Aphorism and Argument," in *Language and Thought in Early Greek Philosophy*, ed. Kevin Robb (Hegeler Institute, 1983).



sought the fundamental constituents of reality and explained change through underlying principles. Aristotle claims that Thales “says the principle is water,” and offers conjectures about why: nourishment and growth depend on moisture, whereby disparate projects become stages in a single investigation.¹²

This has a two-fold negation. First, Aristotle’s narrative arranges earlier thinkers as contributors to a shared investigation and makes their diverse projects look like a continuous enterprise. Second, this retrospective organization introduces a controlled anachronism wherein the Presocratics theorize in Aristotelian terms, yet he interprets them through the language of causes and principles to render them intelligible as predecessors, as inherited by modern scholarship.¹³ As Cherniss shows, reconstructing the Presocratics’ thoughts through Aristotle’s vocabulary filters their ideas through categories that belong to him, not to them.¹⁴

The retrospective construction interacts with another narrative central to the traditional origin story: histories of early Greek thought describe the Presocratics as marking a transition from myth to naturalism. Hesiod’s *Theogony* exemplifies the mythological stage, wherein the origins of the cosmos are a genealogy of divine powers. Chaos comes before the earth and primordial forces, and it is these forces that govern the world. But it is misleading to treat Hesiod as devoid of inquiry; he responds to questions later philosophy will inherit, explaining how the present cosmic order emerged and why it possesses stability rather than chaos. What distinguishes his account from later naturalism is not the absence of explanation but its form: narratives of divine agents and their genealogical relations.

The cultural authority of myth is visible. Herodotus remarks that the Greeks “did not know until yesterday” whence the gods came. Homer and Hesiod composed a theogony, assigned the gods their names, and distributed their honors. Whether accurate, Homeric and Hesiodic poetry shaped the Greek imagination of divine order. Mythic narratives formed the background against which later reinterpretations would emerge.¹⁵

Against this background, the Presocratic shift becomes intelligible. Xenophanes on Iris addresses a phenomenon embedded in

12 Harold Cherniss, *Aristotle’s Criticism of Presocratic Philosophy* (Johns Hopkins Press, 1935); Jonathan Barnes, *The Presocratic Philosophers*, 2nd ed. (Routledge, 1982).

13 Jonathan Barnes, *The Presocratic Philosophers*, 2nd ed. (Routledge, 1982).

14 Harold Cherniss, *Aristotle’s Criticism*.

15 Herodotus, *The Histories*, trans. Robin Waterfield (Oxford University Press, 1998).



Homeric imagery. In the *Iliad*, Iris is Zeus's messenger, conveying divine commands. Xenophanes reinterprets her: "And she whom they call Iris, this too is by nature a cloud."¹⁶ As Leshner notes, Xenophanes' audience would have recognized Iris both as divine messenger and as an atmospheric phenomenon often regarded as a sign from the gods. Xenophanes removes the religious framework and replaces it with natural observation.¹⁷

The transition from myth to naturalism should not be described as a simple replacement of superstition by reason. Xenophanes' reinterpretation presupposes the vocabulary it challenges. "Iris" already carries religious meaning, and his claim gains force by reassigning that meaning within a different framework. What changes is not the existence of inquiry but the standards by which explanations are judged satisfactory.

The same tension appears in figures grouped together by later historians. Heraclitus writes, "Much learning does not teach understanding; otherwise, it would have taught Hesiod and Pythagoras, and again Xenophanes and Hecataeus."¹⁸ The remark functions as a polemical sorting device. Heraclitus acknowledges continuity of inquiry while distinguishing genuine understanding from accumulated information. The grouping reveals that Hesiod represents mythic poetic authority, Pythagoras combines religious and mathematical speculation, Xenophanes criticizes traditional religion, and Hecataeus develops prose accounts that question inherited Greek stories. These figures do not occupy separated domains. They illustrate how boundaries between poetry, theology, cosmology, and emerging critical inquiry remained fluid and contested.

Hecataeus' opening statement makes this explicit. He begins by declaring that he writes what seems true to him because the accounts of the Greeks are many and laughable.¹⁹ This statement does not simply reject earlier narratives. Rather, it introduces a self-conscious stance toward inherited stories. The emphasis falls on evaluating traditional *logoi* according to standards of plausibility and truth. What persists across these figures is, therefore, not a shared doctrine but a

16 Xenophanes, in *Early Greek Philosophy*, Volume III: Early Ionian Thinkers, Part 2, ed. and trans. André Laks and Glenn W. Most, Loeb Classical Library (Harvard University Press, 2016), fragment B32.

17 J.H. Leshner, *Xenophanes of Colophon: Fragments* (University of Toronto Press, 1992).

18 Heraclitus, in *Early Greek Philosophy*, Volume III: Early Ionian Thinkers, Part 2, ed. and trans. André Laks and Glenn W. Most, Loeb Classical Library (Harvard University Press, 2016).

19 Heraclitus, *Early Greek Philosophy*



common space of questions accompanied by evolving criteria for acceptable answers.

Within this shifting intellectual landscape, the emergence of classical philosophy marks an additional development. The activity of philosophy gradually acquires a reflexive self-description. In Plato's *Symposium*, Diotima characterizes the philosopher as occupying a position between wisdom and ignorance.²⁰ The philosopher is neither already wise nor contentedly ignorant but instead desires wisdom and pursues it through inquiry. This description does not yet define philosophy as an institutional discipline, but it provides a conceptual orientation toward a distinctive intellectual activity centered on reasons, understanding, and the pursuit of truth.

Once such self-descriptions stabilize, retrospective construction of philosophical lineage becomes possible. Aristotle presents earlier thinkers as “early philosophers,” and later historians inherit his narrative as if it were neutral discovery. What appears as philosophy's beginning is inseparable from the interpretive frameworks through which later thinkers organize the past. The Presocratics become the first philosophers not simply because of what they said, but because later traditions recognized their inquiries as belonging to the history of philosophy.

IV. GENEALOGY, UNITY, AND REGRESS

Two distinctions help address the relationship between unity and continuity. First, one should distinguish unity of inquiry from unity of discipline. Unity of inquiry concerns whether later thinkers recognize earlier activity as pursuing related questions (cosmos, order, explanation). Unity of discipline concerns whether that activity belongs to “philosophy” under a later description. The standard origin story conflates the two: it takes unity of inquiry as sufficient for unity of discipline. The genealogical view separates them: inquiry can be continuous while disciplinary membership is retrospectively negotiated.

Second, one should distinguish continuity in questions from discontinuity in acceptable explanations. Continuity persists at the level of inquiry; discontinuity lies at the level of accepted explanation. Xenophanes' Iris fragment illustrates: the phenomenon remains; what counts as a satisfying account changes. Divine messaging becomes optional, then suspect, then discarded in favor of physical causes.

²⁰ Plato, *The Symposium*, trans. Robin Waterfield (Oxford University Press, 1994).



One might object: does the acceptance of an explanation determine continuity? Suppose the Presocratics reject Hesiod; does that break continuity or constitute it by keeping the earlier account in view? The genealogical view answers: rejection can be inheritance. Hecataeus treats Greek stories as laughable yet writes in relation to them. Heraclitus criticizes Hesiod because Hesiod holds authority. Xenophanes targets Homeric theology because it matters to his audience. In each case, past inquiry persists as a reference point organizing the present.

This is why “beginnings” are prone to regress. If philosophy begins with the Presocratics because they reject myth for naturalism, we must ask what makes their rejection intelligible as a step in the same story rather than a wholly new practice. The answer cannot be purely doctrinal, because new doctrines break earlier accounts. Instead, the story trades on continuities of question, language, and cultural reference—continuities that extend back through Homer and Hesiod and, by Herodotus’ testimony, through the broader cultural uptake of divine genealogy and naming.²¹

At this point, philosophy’s “beginning” resembles familiar problems of identity over time. We treat an entity as the same across radical change because we maintain a narrative of continuity that selects certain relations of memory, causal connection, and institutional lineage as identity-preserving.²² Likewise, calling the Presocratics “the beginning of philosophy” is less like discovering a birth certificate and more like adopting a principle of narrative organization: we decide the shift in standards of explanation is the identity-preserving thread worth tracking.

This genealogical perspective asks: what kind of story do we want the “history of philosophy” to tell? If we prioritize explanatory standards of naturalistic causes, lawlike order, and criticizable reasons, then the Presocratics become salient. If we prioritize other features (say, philosophy as a way of life, oriented toward self-transformation and spiritual exercise), then the story shifts toward later schools and practices, as emphasized by Pierre Hadot’s influential thesis that ancient philosophy is better understood primarily as a way

21 On the relationship between myth and early philosophical thought, see Jean-Pierre Vernant, *The Origins of Greek Thought* (Cornell University Press, 1982); and G.E.R. Lloyd, *Demystifying Mentalities* (Cambridge University Press, 1990).

22 On personal identity and continuity across change, see Derek Parfit, *Reasons and Persons* (Oxford University Press, 1984); and Marya Schechtman, *The Constitution of Selves* (Cornell University Press, 1996).



of life rather than as doctrine alone.²³ The “beginning,” on the genealogical view, is therefore not merely found; it is argued for.

V. OBJECTIONS AND REPLIES

Perhaps the meaning of “philosophy” changes, but the practice itself retains a stable essence. On this view, the question of origin remains meaningful because it asks when that activity first appeared. The “natural kinds” analogy is often invoked. Water’s meaning changed, but the referent remained stable because the underlying substance was always H₂O. One might suggest philosophy is similar: its definition evolves, but the activity has a stable nature gradually recognized.

The difficulty is that philosophy does not resemble a natural kind. Putnam and Kripke argue that reference is fixed by an external structure independent of human practices. Water has a chemical composition that determines what counts as water regardless of how speakers describe it. Philosophy, by contrast, exists only through human practices.²⁴ The historical record shows that the extension of *philosophia* was shaped by institutions, rhetorical conventions, and disciplinary boundaries. Early uses refer broadly to intellectual cultivation; later uses designate a more specialized inquiry.²⁵ This resembles the gradual stabilization of a practice, not the discovery of a hidden essence.

A second objection appeals to the rupture between myth and naturalistic inquiry. The transition associated with early Greek thinkers, especially Xenophanes, is sometimes treated as a decisive turning point: explanations based on divine agency are replaced by accounts grounded in natural principles. Xenophanes’ claim that Iris is a cloud exemplifies this shift. If this represents a genuine intellectual revolution, philosophy begins when explanations first appeal to internal natural principles.

This captures something important, yet it depends on retrospective interpretation. Xenophanes’ reinterpretation is intelligible as a challenge only where Iris already possesses mythic significance. The shift presupposes the conceptual vocabulary inherited from earlier traditions. Even Aristotle complicates the picture. When discussing Thales, he remarks that earlier mythic accounts involving Ocean and

23 Pierre Hadot, *Philosophy as a Way of Life*.

24 Hilary Putnam, “The Meaning of ‘Meaning.’”

25 On the semantic range of *philosophia*, see Pierre Hadot, *Philosophy as a Way of Life*; and Alexander Nehamas, *The Art of Living* (University of California Press, 1998).



Tethys may contain ideas that resemble naturalistic principles.²⁶ By framing history this way, Aristotle blurs the boundary between myth and nature and renders earlier narratives interpretable as proto-naturalistic. What appears as a clear beginning depends on later reinterpretation.

A third objection concerns acceptable explanations. If the argument relies on changes in explanatory standards, can such standards mark discontinuity? Explanations are rarely final. They remain open to revision. If so, it seems difficult to identify any clear point where one form genuinely replaces another. But discontinuity concerns justificatory norms, not logical possibility. Earlier explanations do not become impossible. They cease to function as persuasive within a particular intellectual community. Iris as divine messenger is not refuted like an incorrect measurement. It loses its status as acceptable once new norms emphasizing observation and natural causation gain authority. The shift concerns the standards by which explanations are judged, not the metaphysical impossibility of earlier accounts.

One might worry that the genealogical view makes philosophy's beginning arbitrary. If origins are retrospective constructions, couldn't any starting point be chosen? But the genealogical approach does not hold that all origin stories are equally valid. Rather, historical narratives must be evaluated by the explanatory work they perform. Different accounts highlight different features of intellectual development. Some emphasize the emergence of naturalistic explanation; others stress philosophy as a way of life or ethical reflection.²⁷ The genealogical view does not abandon the search for beginnings. It reframes the task: beginnings depend on interpretive choices about which forms of continuity we regard as most significant.

VI. CONTEMPORARY IMPLICATIONS

If philosophy's beginning is a retrospective judgement about what counts as part of "our" lineage, contemporary philosophy inherits two responsibilities: clarify the standards for inclusion and examine the consequences of the historical narratives through which the discipline represents its past.

26 On Aristotle's doxographical method, see Harold Cherniss, *Aristotle's Criticism*.

27 For the "way of life" interpretation, see Hadot, *Philosophy as a Way of Life*. For the naturalistic emphasis, see G.E.R. Lloyd, *Magic, Reason and Experience* (Cambridge University Press, 1979).



The first implication concerns method. The genealogical view supports the methodological restraint emphasized by contextualist approaches.²⁸ Interpreters should not assume philosophical problems remain identical across time or treat earlier texts as contributors to modern debates. Rather, interpretation must reconstruct what could plausibly be meant within the conceptual environment where a text was produced. This does not prohibit engagement with earlier thinkers. It requires distinguishing two conflated activities: historical reconstruction and philosophical appropriation. The first seeks to understand what earlier authors were doing in their own context. The second asks how those texts might contribute to contemporary inquiry.

A second implication concerns the narrative that philosophy emerges by replacing myth with rational explanation. The analysis of the myth–nature transition suggests this is too simple. Many texts that introduce naturalistic explanation remain connected to earlier frameworks. Xenophanes criticizes traditional gods while developing his own theology. Heraclitus attacks polymaths but appeals to a logos that structures the cosmos. Aristotle treats mythic narratives as containing rudimentary insights about nature.²⁹ These examples suggest the shift from myth to naturalism is not a clean break. The boundary between myth and philosophy emerges gradually through changing standards of explanation.

Recognizing this complexity changes how we understand philosophical unity. Unity does not require a single method, set of problems, or origin. It may consist in an ongoing practice of reinterpretation through which the discipline revises its boundaries and reconsiders what belongs within its tradition. Historical narratives shape how philosophers identify predecessors and organize conceptual inheritance.

Here, Pierre Hadot is instructive. He argued that many ancient schools understood philosophy not primarily as doctrines but as a way of life structured by intellectual and ethical exercises.³⁰ If correct, the history of philosophy contains multiple conceptions of what the activity consists of. Such plurality makes it unlikely that any single origin story captures the full range of philosophical practice.

These considerations point to a final implication. If philosophy begins when it becomes possible to reflect on its own beginning, metaphilosophy is not optional but internal to inquiry. Questions

28 On contextualism in intellectual history, see Quentin Skinner, "Meaning and Understanding in the History of Ideas," *History and Theory* 8, no. 1 (1969).

29 Aristotle, *Metaphysics*; Harold Cherniss, *Aristotle's Criticism*; Jonathan Barnes, *The Presocratic Philosophers*.

30 Pierre Hadot, *Philosophy as a Way of Life*.



about origin reveal the standards by which the discipline recognizes itself. They bring into view what counts as explanation, whose voices count as predecessors, and which forms of discourse are excluded as myth, poetry, or religion. The question of philosophy's beginning belongs not only to historical investigation but to philosophy's continuing effort to interpret its own activity.

VI. CONCLUSION

The familiar answer that philosophy begins with the Presocratics is defensible relative to a particular standard of explanation: these thinkers mark the beginning because their accounts appeal to principles internal to nature, relying on observation, criticism, and argument rather than divine genealogy. Yet this criterion is historically situated. It becomes authoritative only through retrospective interpretation, most clearly in Aristotle's organization of earlier thinkers into a unified search for causes and principles.

Once we distinguish continuity of inquiry from continuity of disciplinary identity and recognize that rejecting earlier explanations can constitute inheritance, the idea of an absolute beginning becomes difficult to sustain. Hecataeus on Greek logoi, Xenophanes on Iris, and Heraclitus on earlier authorities show that later inquiry can remain continuous with traditions it rejects. What persists is not a stable set of doctrines but a field of questions and practices through which earlier accounts continue to shape later reflection.

Philosophical origins are, therefore, best understood genealogically. Philosophy begins when later thinkers recognize earlier inquiry as part of their own past. The beginning is not a moment to be discovered but a judgement about how certain practices come to be described. Philosophy begins whenever inquiry turns back and claims a lineage.







Aarush Atul Prabhu (he/him) is a law student with primary interests in philosophy and sociology. After completing his undergraduate program, he intends to develop a unified theory of how knowledge, identity, and social systems evolve. Immense thanks to family, Rishi and Jiya for their support.



LEARNING TO LIVE “THE REAL THING”

Contemplation, *Acedia*, and
Redemption in *The Death of Ivan Ilyich*

ABSTRACT

This paper argues that Leo Tolstoy’s *The Death of Ivan Ilyich* presents not merely a cautionary tale, but an account of spiritual reorientation and redemption. Bringing Josef Pieper’s concepts of *acedia* and contemplation into dialogue with Tolstoy, it contends that Ivan’s confrontation with death exposes the misalignment of his life while opening the possibility for transformation. Ultimately, Ivan’s final moments reveal that authentic human flourishing lies not in social success, but in contemplative openness to reality and a redemptive consent to truth—even at life’s end.

ALEXANDER ROTH



I. INTRODUCTION

When I first read *The Death of Ivan Ilyich*, I was struck, and almost unsettled, by how depleted, hollow, and tragically ordinary Ivan's life seemed. Yet when I reread Tolstoy's novella, I came to realize that Ivan is not merely a cautionary symbol of a life gone wrong; in fact, quite the opposite: Ivan actually serves as an exemplar of what it means to reorient one's life in the face of death.

What initially seemed like a cautionary tale of squandered potential, *The Death of Ivan Ilyich* provides a transformative vision grounded in contemplation and spiritual reorientation. As Ivan lives out his final days, squirming in anguish from his terminal sickness, he reconciles his shortcomings, turning away from the dark abyss that previously consumed his life and toward the bright light of authentic love. This moment of conversion marks a true metanoia:¹ far from a denial of suffering, his transformation is a march through it, allowing death to strip away his vanity so that he may love without fear.

While many scholarly interpretations treat Ivan as a tragic warning against a superficial bourgeois life, this paper will claim that Tolstoy presents Ivan as an unexpected exemplar of spiritual reorientation. Although his early life is marked by distraction and *acedia*, Ivan's final contemplative consent to reality reveals that redemption remains possible until the final breath is drawn. To defend this claim, this paper proceeds in three steps: first, it examines how Ivan's confrontation with death exposes the misalignment of his life; second, it brings Josef Pieper's account of *acedia* and contemplation into conversation with Tolstoy; finally, it argues that Ivan's death is not merely tragic but redemptive, revealing that human flourishing consists not in success or control, but in consent to reality and openness to what exceeds us.

II. WHAT DOES IT MEAN TO LIVE “THE REAL THING”?

First, let us explore, in more depth, the question that sits at the heart of Leo Tolstoy's *The Death of Ivan Ilyich*: What does it mean to live “the real thing?”

For most of his life, Ivan orients himself around what is socially approved rather than what is inwardly true. He strives to become a

1 The word ‘metanoia’ comes from the Ancient Greek ‘*meta*’ (changed) and ‘*nous*’ (mind), meaning ‘a change of mind,’ or ‘a fundamental change of heart.’



well-respected legal official, cultivates proper manners, and structures his family life according to convention, mistaking external success for genuine fulfillment. His days are marked by efficiency and respectability, yet largely devoid of reflection, and moral depth. Thus, as his illness worsens, Ivan is forced into a painful moral clarity: the possibility that he has not truly lived at all. In a moment of severe physical pain, and immense psychological torment, he reflects, “But no matter. I can still make it the real thing—I can. But what is the real thing?”² His concluding question exemplifies the overarching tension of the human condition: the fear that one’s life may have been inauthentic, misdirected, or un-lived.

Yet even this fear contains an unexpected grace. As the Roman philosopher Cicero once proclaimed, *dumanimaest spesesse*, or “While there is life, there is hope.”³ It is never too late to turn toward virtue and to confront death—not as a final terror, but as a mirror revealing how we have truly lived. Josef Peiper, in *Leisure, the Basis of Culture*, offers a philosophical framework for answering the very question that torments Ivan: How can one reorient oneself towards “the real thing?” In short, both Tolstoy and Peiper would argue that we must remain open to what exceeds us⁴—open, that is, to the moral and metaphysical realities that lie beyond the self. These works reveal that death does not simply end life—it exposes life’s authentic orientation, disclosing whether it was marked by distraction or directed toward the Truth. Dying well, then, is only possible if one has lived well; yet, until the last breath is drawn, there is always the potential for reorientation toward reality and toward that which transcends human limitation.

Ivan Ilyich, despite a life initially marked by distraction and carelessness, ultimately came to live “the real thing” in the final moments before his death. Ivan’s life can be aptly framed through a reflection from Shelly Kagan’s *Death*:

Recall Tolstoy’s *Death of Ivan Ilyich*. The people in the Tolstoy story seem to have put the fact of mortality out of their mind. Why? Presumably because they think that facing it is just too crushing and overwhelming. So the way they cope with it is by simply disregarding the facts, trying not to think about them.⁵

2 Tolstoy, Leo. *The Death of Ivan Ilyich* (Bantam Classics, 1981), 71.

3 Cicero, *Letters to Atticus*, Book 9, Letter 10.

4 The phrase “we must be open to what exceeds us” is a Pieperian coined term, meaning that we must look beyond the mundane and acknowledge something greater than ourselves—namely, transcendence or the divine reality. This clarification is important because it establishes the philosophical grounding of the term, preventing it from being misinterpreted, and situating the argument within Pieper’s account of transcendence as essential to human flourishing.

5 Kagan, Shelly. *Death* (Yale University Press, 2012).



This description captures not only the society that shapes Ivan, but also Ivan himself—an individual who spent much of his life evading the essential truths that prompt self-examination. His pursuit of comfort and status served as diversions from confronting the finitude that quietly shaped his existence. Ivan may have been aware of death, but he never allowed that awareness to shape how he lived—as if mortality were always a distant concern. In refusing to live with *memento mori*, the honest remembrance of death, Ivan’s pursuit of societal recognition becomes a lifelong attempt to escape the very truths that give life depth.⁶ It is only when these illusions collapse that he begins to recognize the essence of a life authentically lived.

III. DEATH AND THE COLLAPSE OF ILLUSION

Nonetheless, Ivan’s journey affirms the central claim that until the last breath is drawn, the potential for reorientation toward reality remains. Ivan’s illness, while outwardly agonizing, births redemption, serving as a site for profound spiritual transformation. Upon first encountering his diagnosis, Ivan retreats behind psychological “screens” to defend himself from the looming reality of death. As Tolstoy narrates, “to save himself from this condition Ivan Ilyich looked for consolations—new screens—and new screens were found and for a while seemed to save him, but then they immediately fell to pieces or rather become transparent, as if It penetrated them and nothing could veil It.”⁷ This metaphor of “veils breaking down” illustrates the collapse of illusion in the face of ultimate Truth. Tolstoy not only exposes Ivan’s fear of suffering, but his resistance to viewing his life as it truly is. The “screens” that Ivan fascinates himself with function as forms of self-deception, preserving a way of life structured around distraction rather than honesty. Thus, once these defenses fail, Ivan is left with no refuge except reality itself. It is precisely this forced encounter with truth that makes genuine reorientation possible, revealing that redemption does not occur despite suffering, but through the clarity suffering imposes.

Tolstoy shows that the true catalyst for Ivan’s crisis is not the physical pain, but the impossible-to-ignore fact of his quickly approaching death. As Perrett explains,

First, it is the knowledge of death that once again generates the crisis. Ivan (like any adult person) must know that he will

6 “Memento mori” is a Latin phrase that translates to “remember you must die,” serving as a philosophical reminder of the inevitability of death used to encourage living a more purposeful, virtuous, and meaningful life.

7 Tolstoy, Leo. *The Death*, 51.



die, but since he refuses to face this fact, he has no idea of how to live in the light of it. He does not know how to live. In the face of his death he comes to know that the way he has lived is wrong, for his death renders meaningless the life he has led.⁸

Ivan's frantic need for distraction—his desperate attempts to maintain a sense of control—only highlights how thoroughly he had built his life on evasion. Once death punctures every barrier, the shallow interior of his life is exposed.

Ivan's frantic need for distractions to maintain his sense of control becomes increasingly inadequate as death punctures every barrier. At one decisive moment, Tolstoy shows Ivan brushing up against the very insight that could transform him: "‘Maybe I did not live as I ought to have done,’ it suddenly occurred to him. ‘But how could that be, when I did everything properly?’ he replied, and immediately dismissed from his mind this, the sole solution of all the riddles of life and death, as something quite impossible."⁹ Even when clarity emerges, Ivan initially refuses it—clinging instead to the life he had curated with such care. This moment reveals the tragic logic that governs Ivan's self-understanding: the assumption that one's reputation is equivalent to moral truth. Ivan cannot yet conceive that a life "properly" lived could be fundamentally misaligned with reality. His refusal comes not from ignorance, but from a defense of identity. If Ivan were to accept this insight, his life would collapse, as he would be exposed to the fundamental flaw in the very standards by which he has measured his worth. Thus, Tolstoy suggests that the greatest obstacle to reorientation is not the absence of insight, but the cost of accepting it: the surrender of a self built upon comfort, order, and social approval.

IV. IVAN'S ORDINARY LIFE

Tolstoy uses this contemplative inquiry to underscore the role of submission in viewing reality. Also, as several scholars note, Tolstoy never intended Ivan's situation to be viewed as an isolated tragedy, but instead a rather common one. Perrett successfully captures this insight: "Although it is Ivan's life that the story presents for us, it is also quite clear that his case is to be viewed as an entirely typical one: 'Ivan Ilych's life had been most simple and most ordinary and therefore most terrible.'"¹⁰ This observation underscores the unsettling universality of

8 Perrett, Roy, "Tolstoy, Death and the Meaning of Life," *Religious Studies* 26, no. 2 (1990): 241, <https://www.jstor.org/stable/3751001>.

9 Tolstoy, Leo. *The Death*, 64.

10 Perrett, Roy, "Tolstoy, Death," 238.



Ivan's situation. His life is "terrible" not because it is monstrous, but because it is painfully ordinary—structured around social approval, comfort, and routine rather than reflection, virtue, or genuine love. Tolstoy's warning is clear: ordinary life, when unexamined, becomes a quietly catastrophic one. Thus, Ivan's story is not meant to isolate him as a moral outlier, but to warn the reader of how easily a socially successful life can drift into spiritual emptiness without ever appearing outwardly disordered. In this way, Tolstoy transforms Ivan from a singular character into a mirror, urging the reader to reflect on the dangers of societal norms. However, Ivan's suffering becomes a redemptive blaze, purging him of pride and forcing him to confront the question that lies beneath all human striving: "How. . . then. . . shall [I] live?"¹¹

V. SUFFERING, THE TURN INWARD

As Ivan's death approaches, he turns inward and begins to explicitly wrestle with the meaning of his suffering in relation to God. In one of the most theologically charged passages of the novella, Ivan cries out, "Why hast Thou done all this? Why hast Thou brought me here? Why, why dost Thou torment me so terribly?"¹² Initially, this appears to be an outburst of exasperation, but Tolstoy quickly shifts the tone. Ivan "grows silent," holding his breath and becoming attentive not to an external voice, but to the stirrings of his own soul: "It was as though he were listening not to an audible voice but to the voice of his soul, to the current of thoughts arising within him . . . 'What do I want? To live and not to suffer.'"¹³ This moment of reflection reveals Ivan's desire for a life without fear—a life aligned with God. The response that rises from within, "As you lived before, well and pleasantly?"¹⁴ is piercing. The question mark signals Ivan's first genuine hesitation of the standards by which he once measured a good life.

Through this inquiry, Tolstoy exposes the hollowness of a "happy life" rooted in ease rather than truth. What once appeared societally acceptable now appears foul, underscoring the contradiction between comfort and virtue that Ivan can no longer ignore. This question confronts Ivan with the truth of his life's orientation: that he had loved comfort more than virtue, and that he had loved himself more than the Truth. The role of stillness in contemplation proves to be essential,

¹¹ Tolstoy, Leo. *The Death*, 71.

¹² Tolstoy, Leo. *The Death*, 70.

¹³ Tolstoy, Leo. *The Death*, 70.

¹⁴ Tolstoy, Leo. *The Death*, 71.



serving as a quiet opening of oneself to the reality that exceeds them. In allowing himself to hear the voice of his conscience rather than fear isolation, he begins the slow work of reorienting his soul toward the truth.

VI. IVAN'S FINAL GESTURE: A STRETCH TO THE DIVINE

In a final effort to repair the neglect he had constantly shown his family, and as his wife and children gather beside him, Ivan attempts to say, "Forgive me," but can only utter "forego," while waving his hand—"knowing that He whose understanding mattered would understand."¹⁵ In this culminating moment of vulnerability, Ivan sheds his former self and opens his entire being to the divine. The beauty of this closing moment lies not in his action but in the humility of his gesture, a final submission to the truth. By explicitly naming God in his final moments, Ivan opens himself to what exceeds him and deliberately decides to love God more than himself. Thus, Ivan's death becomes something truly beautiful. In the final moments of his life, he finally embraces the gift he had been given and turns toward the reality that transcends him. In the end, it is death itself that enables Ivan to become fully human via the act of contemplation.

VII. ACEDIA: THE HIDDEN DISORDER OF IVAN'S LIFE

By bringing Josef Pieper into conversation with Tolstoy, two insights come into focus. First, Pieper's account of *acedia* provides a name for the spiritual disorder that quietly shapes Ivan's misguided life long before his illness. Second, Pieper's understanding of contemplation illuminates how Ivan, at the very end of his life, is finally able to turn toward the truth he had spent so long avoiding.

Pieper's diagnosis of Ivan's life can be captured by a single term first articulated by the fourth-century monastic thinker Evagrius Ponticus. *Acedia* offers a useful lens for understanding the spiritual emptiness that shapes Ivan's life long before he ever becomes physically ill. Pieper defines *acedia* as follows:

Acedia is the 'despair from weakness' which Kierkegaard analyzed as the 'despairing refusal to be oneself.' Meta-physically and theologically, the notion of *acedia* means

¹⁵ Tolstoy, Leo. *The Death*, 82.



that a man does not, in the last resort, give the consent of his will to his own being; that beneath the dynamic activity of his existence, he is still not at one with himself.¹⁶

This diagnosis describes Ivan's condition with utter precision. Throughout the novella, Ivan avoids the deeper demands of selfhood by immersing himself in the distractions of career, propriety, and social status. Pieper captures this dynamic with a single line: "A man who cannot be still cannot perceive the truth of the world."¹⁷ Ivan's overwhelming busyness masks an interior refusal to confront who he truly is and what his life is ordered toward. Pieper warns that such restlessness is not merely personal but culturally reinforced within what he calls the totalitarian world of work—a social order that measures human worth almost exclusively by productivity and outward success.¹⁸ Ivan's life unfolds almost entirely within this framework, leaving little space for contemplation, leisure, or genuine self-encounter. In Pieperian terms, Ivan's lifelong pursuit of comfort is a symptom of spiritual negligence, as he refuses to affirm the truth of his own being. His later transformation, then, becomes clear when viewed in light of his underlying condition, which Pieper would define as *acedia*.

VIII. CONTEMPLATION AS REORIENTATION

If *acedia* names the sickness that shapes Ivan's life, contemplation becomes the saving grace that finally halts it. For Pieper, contemplation is not an act of grasping or analyzing, but a posture of receptive openness: "To contemplate, on the other hand, to 'look' in this sense, means to open one's eyes receptively to whatever offers itself to one's vision, and the things seen enter into us, so to speak, without calling for any effort or strain on our part to possess them."¹⁹ This is the very practice that Ivan had avoided for most of his life. Instead of reflecting on the world around him, Ivan filled his days, fleeing from anything that might disrupt his carefully curated comfort. Yet contemplation demands the opposite: an unguarded willingness to receive reality as it is. As Pieper puts it, "The person who contemplates is one who consents to the world, who says 'yes' to the whole of creation."²⁰ Only in his final days when his illusions crumble and he can no longer distract himself

16 Pieper, Josef. *Leisure: The Basis of Culture* (Ignatius Press, 2009), 44.

17 Pieper, Josef. *Leisure: The Basis of Culture*, 41.

18 Pieper's "totalitarian world of work" resembles a culture that prioritizes utilitarian production, making work the center of existence which spiritually impoverishes people.

19 Pieper, Josef, *Leisure*, 26.

20 Pieper, Josef, *Leisure*, 37.



does Ivan begin to practice this kind of receptivity. Ivan finally opens himself, in painful stillness, to the truth of his existence, and to the needs of those around him. Ivan discovers that redemption does not come through correction, or achievement, but through a fundamental reorientation of vision—a willingness to affirm reality rather than flee from it. Thus, Ivan’s final contemplative act becomes the very root of his transformation, revealing that it is not the avoidance of suffering but the consent to reality, even when it hurts, that allows a human life to be set right.

IX. CONCLUSION: DEATH AS THE ULTIMATE TRUTH

In short, this reflection underscores the central role of contemplation in everyday human life and the necessity of remaining open to what exceeds us. Pieper reminds us that true flourishing, contrary to today’s materialistic implications, begins when we stop grasping at life, and simply receive it when we submit to reality, rather than resist it. Ivan Ilyich, in his final days of life, serves as an exemplar of the truth of Pieper’s message. His life, so long shaped by *acedia* and avoidance, becomes a case study for the transformative power of contemplation.

Tolstoy’s portrayal of Ivan ultimately advances a decisive claim: a human life is not measured by its material success, but by its orientation toward reality. When read alongside Pieper’s account of contemplation, Ivan’s final transformation reveals that even a misdirected past does not close the door to redemption. A life shaped by distraction and *acedia* can, in its final moments, be reoriented toward truth. Ivan Ilyich is therefore not merely a cautionary tale, but an exemplar of how suffering, rightly received, can become the site of spiritual renewal.





Alexander R. Roth (he/him) is a junior at Villanova University, majoring in humanities, philosophy, and theology, with minors in creative writing and ethics. His interests include C.S. Lewis, Thomistic theology, the problem of evil, and existential themes in Lewis and Tolstoy. He writes on contemplation, suffering, and moral transformation.

A STOIC CASE FOR THE THE LAND ETHIC

ABSTRACT

In this paper, I aim to show the intersection of conventional Stoic thought with Aldo Leopold's Land Ethic, which he introduces in *A Sand County Almanac*. I first explore the concept of a 'preferred indifferent.' Diogenes Laërtius, paraphrasing the early Stoics, describes it as "a certain intermediate potential or usefulness which contributes to the life according to nature." I then argue that the "land," as Leopold conceives it, is one such preferred indifferent, and that taking action to maintain it is therefore an expression of Stoic virtue befitting of the "life in accord."

JONAH BLACK



I. INTRODUCTION

Alfred North Whitehead is famous for his description of modern Western Philosophy as “footnotes to Plato.” In this treatise, I explore the interplay between Stoicism and Leopoldian Naturalism to illustrate how these distinct schools converge on similar conclusions about the significance of land. I first analyze the Stoic notion of preferred indifferents, which, though not virtuous in and of themselves, can aid an individual in pursuing virtue. Then I turn to Stoic conceptions of the cosmos and how their ideas of divine will and cosmic order are echoed in modern understandings of natural structures. I then explore the concept of natural equilibria through the story of the Aral Sea. Though not a consequentialist philosophy, Stoic doctrine holds that natural equilibria ought to be preserved, and the aftermath of the Aral Sea catastrophe illustrates the cost of their disruption. Finally, I analyze Aldo Leopold’s “Land Ethic,” which he introduces in *A Sand County Almanac*, and discuss how it fits into the Stoic tradition. I situate land among the preferred indifferents and thereby show that actions promoting the “goodness” of the land are virtuous in the broader Stoic account, insofar as they enable a state in which the persons residing within can more easily seek virtue. This, in turn, reinforces the notion that our contemporary concern for the environment is, to borrow from Whitehead, a footnote to ancient Stoic thought.

II. PREFERRED INDIFFERENTS AND THE NATURAL WORLD

The Stoics hold that virtue derives from the “life according to nature,” and that achieving this requires aligning oneself with the natural order by exercising reason; thus, virtue is accessible only to rational beings. Moreover, while external things can affect one’s ability to achieve virtue, the Stoics consider them “indifferent,” meaning they are neither good nor bad in themselves. Indeed, to preoccupy oneself with such things is inherently in tension with the Stoic way of life. This sentiment is found in such dictums as Epictetus’s “It is not events that disturb people, it is their judgements concerning them,”¹ Marcus Aurelius’s “Look within; within is the fountain of all that is good,”² and Seneca’s “...that virtue is the sole good, and certainly that nothing is

1 Epictetus, *The Enchiridion*, trans. Robert Dobbin (Penguin Books, 2008), 29.

2 Marcus Aurelius, *Meditations*, trans. Robin Hard, ed. Christopher Gill (Oxford University Press, 2016), 7.31.



good without virtue; moreover, that virtue itself is located in our better part, namely, the rational part.”³ Although differing in expression, these statements reflect the Stoic notion that “living in accord” is a fundamentally internal phenomenon that only manifests in action.

Within this framework, preferred indifferents are things that, while not inherently good, can aid an individual in their pursuit of virtue. Diogenes Laërtius, paraphrasing the early Stoic philosophers, describes them as “a certain intermediate potential or usefulness which contributes to the life according to nature.”⁴ He further notes the Stoics’ use of the word “good” to refer to “that from which there is something beneficial.”⁵ In this context, “good” is an anthropocentric term used to describe things that further the human pursuit of virtue. However, there is a semantic difference between the descriptive “good” and the prescriptive “the good,” the latter of which refers to “life in accord with nature,” which is the highest form of human existence in the Stoic tradition—a state achieved by the “sage.”

Stoic teachings about external things do not preclude the possibility of rational action involving them, provided such actions are in accord with virtue. If we recognize the land as a preferred indifferent, then the maintenance and care of the land enters the scope of virtuous action. Environmental stewardship, insofar as it entails rational action in accordance with nature, therefore becomes an expression of Stoic virtue. This bridge between disposition and conduct enables the Stoic to explain why we engage with the external world, while still conceiving of virtue as a fundamentally internal state.

For the Stoics, what it means to live virtuously, or “in accord with nature,” varies slightly by account. Diogenes, paraphrasing the views of the early Stoic philosophers such as Chrysippus and Cleanthes, defines Stoic virtue as “that which is perfectly in accord with nature for a rational being, qua rational.”⁶ The consensus among early Stoics is that to live in accord with nature is to live in accord with prudence, which he defines as “the knowledge of which things are good and bad and neither.”⁷ Cicero, a notable Stoic-adjacent thinker, makes a slightly stronger claim. He argues that, concerning the “cosmos” and everything it contains, “a good and wise man, who is familiar with the laws and is not ignorant of civil duty, consults the welfare of all more than

3 Seneca, *Letters to Lucilius*, trans. E. Phillips Barker (Clarendon Press, 1932), 71.32.

4 Diogenes Laërtius, *Lives of Eminent Philosophers*, ed. Tiziano Dorandi (Cambridge University Press, 2013), 7.105.

5 Diogenes Laërtius, *Lives of Eminent Philosophers*, 7.94.

6 Diogenes Laërtius, *Lives of Eminent Philosophers*, 7.94.

7 Diogenes Laërtius, *Lives of Eminent Philosophers*, 7.92



that of one person or his own.”⁸ Notable here is Cicero’s use of the words “laws” and “civic duty” in the context of the cosmos as a whole, which is consistent with his subsequent claim, originally articulated by Chrysippus, that “all other things [besides men and gods] were created for the sake of men and gods.”⁹ This view sets Cicero apart from many of the Stoics in that he directly acknowledges the importance of the cosmos’s particular structure and intricacies.

The mechanisms of the cosmos, whose function and purpose are not always apparent, work together to produce an environment in which humans (and, in the ancient account, gods) can thrive. For humans, who lack omniscience, it is difficult to ascertain exactly what effects these mechanisms have on us. The prudent course of action is to preserve these mechanisms as best we can, since, due to the Stoic belief in their divine provenance, they may further our well-being in ways that are not immediately obvious. As prudence is a cardinal Stoic virtue, this course of action aligns with the “life in accord,” bringing one closer to the status of sage.

Marcus Aurelius takes a similar view: “That rational essence by which the universe is governed is for community and society.”¹⁰ While we may nowadays question the notion of intelligent design, or that our universe is the product of a divine will, we might still ask: Does rationality necessarily imply intention? The dynamic systems that promote our survival in the absence of conscious intent operate in a manner that appears rational, in the sense that we exist to observe them. Seemingly, then, we are indeed governed by such an essence—perhaps not a singular or even conscious one, but one that accords with the Stoic understanding of the natural order, and which is reflected in the principle of *logos*. *Logos*, for the Stoics, is the active principle that describes the structure and behavior of the universe, and it is to this active principle that we ascribe the term “essence.” The precise character of this essence varies slightly by account; for our purposes, we may refer to this essence as the laws of nature. These natural laws operate systematically, in accord with a set of immutable principles that define the scope of possible occurrences. While perhaps not rational in the human sense of the term, the land itself and the qualities it affords us are included in this scope and contribute directly to our “community and society.” Our relationship with the land, which is shaped by environmental conditions, is in this way an expression of a rational natural order. When land exists in an equilibrium in which

8 Cicero, *De Finibus Bonorum et Malorum*, trans. H. Rackham (Harvard University Press, 2006), 3.64.

9 Cicero, *De Finibus Bonorum et Malorum*, 3.67.

10 Marcus Aurelius, *Meditations*, 5.24.



life can prosper, we might call it *good*. Such land is a preferred indifferent because its existence furthers our livelihood as rational beings and thus facilitates virtuous action. Just as health or wealth can enable an individual to engage more effectively in virtuous activities, so too can good land enable communities and individuals to thrive and cultivate virtue.

III. THE LESSON OF THE ARAL SEA

Having established that actions concerning external things can be consistent with Stoic virtue, we now turn to a more concrete example that demonstrates the importance of environmental stewardship: the Aral Sea. Through this example, I aim to illustrate both how land contributes to our continued existence in a stable state and the consequences of excessive tampering with natural structures.

For a long time, the Aral Sea was one of the largest lakes in the world. Many local communities relied on it as the staple of their livelihood, well into the 1950s. In the late twentieth century, Soviet Russia, seeking to grow cotton in the region, began a series of irrigation projects to divert the Aral Sea's tributaries into the surrounding desert. As a result, the sea began to dry up. Communities reliant on fishing and agriculture were permanently affected, and the exposed seabed released pollutants into the air, causing health problems for local populations. The environmental effects were severe and widespread, tearing at the region's socioeconomic fabric. This area of the world has yet to recover from this catastrophe, and the damage done is a direct consequence of our attempting to assume a godlike role that exceeds our understanding.

Standard modern practice for assessing an event like this is to put it in terms of expected cost-benefit, or to use a similar consequentialist method. In fact, the Stoic perspective, although *pre hoc* and outcome-agnostic, is equally instructive. Recall that prudence, as the Stoics conceived it, is "the knowledge of which things are good and bad and neither." The goodness of an action, then, cannot derive from its output, even in cases where its output is desirable. Epictetus further counsels us to, "reflect on what every project entails in both its initial and subsequent stages before taking it up."¹¹ Prudence thus requires us to acknowledge the limits of our understanding and not alter what we have determined is preferred without due consideration of what the alteration entails. Had the Aral Sea project succeeded, then, the fact that

11 Epictetus, *The Enchiridion*, 29.1



it required tampering with a nominally good thing would have made it imprudent in the Stoic account, hence, not virtuous. The frameworks of Aurelius and Cicero likewise support the view that the Soviets, in not consulting the “welfare of all” and neglecting to consider the land’s contribution to “community and society,” failed to meet the standards for goodness articulated in the two accounts. Thus, whereas a consequentialist account considers the outcome, the Stoic one considers the goodness of the *reasoning* behind the action. In this way, virtue, as it is classically understood, can be preserved.

The Aral Sea is a testament to the era-agnostic human tension between hubris and fallibility. Leopold warns us against this in his *Almanac*: “In human history, we have learned (I hope) that the conqueror role is eventually self-defeating. Why? Because it is implicit in such a role that the conqueror knows, *ex cathedra*, just what makes the community clock tick, and just what and who is valuable, and what and who is worth less, in community life.”¹² This quote embodies a core tenet of the Land Ethic: to act as a conqueror of nature is to misunderstand our place within it. Like the Stoics, Leopold is wary of human fallibility and therefore careful in making normative claims about the natural order. Both parties understand that human attempts to categorize natural things according to their impact on well-being are prone to errors, both because of our limited grasp of nature’s mechanisms and because we tend to overestimate ourselves in this respect. Therefore, according to Leopold, “a thing is right when it tends to preserve the integrity, stability, and beauty of the biotic community. It is wrong when it tends otherwise.”¹³ His reasoning follows from the principle that there exists a strong, interdependent relationship between individuals and the land upon which they reside. Land, he argues, enables communities to come together and thrive, and the community, in turn, allows individuals to flourish. As such, our relationship with the land should be underpinned by an ethical framework that explicitly accounts for it. The land’s well-being is the community’s well-being and, by extension, the individual’s.

IV. FINDING VIRTUE IN RECIPROCAL PRESERVATION

The notion of reciprocal preservation outlined here is reminiscent of Stoic ideals first articulated two millennia ago. In Cicero’s words, “Every animal, as soon as it is born, has an affinity to itself

12 Aldo Leopold, *A Sand County Almanac*, ed. Kurt Brower and Michael Sewell (Oxford University Press, 2001), 193.

13 Aldo Leopold, *A Sand County Almanac*, 211.



and is inclined to preserve itself and its constitution and to like those things which preserve that constitution.”¹⁴ It follows, therefore, that a rational person, recognizing that the flourishing of the land is inextricably linked with their own, would be disposed to care for the land as a means to preserving their own constitution. Failure to do so would amount to a form of self-sabotage, since it would undermine their own well-being. Thus, prosperous land supports Cicero’s notion of self-preservation, which in turn engenders more opportunities to pursue virtue. In this sense, land exemplifies the primary characteristics of a preferred indifferent.

Moreover, among the preferred indifferents, land may be the most foundational, given its potential to impact the accessibility of others. Harsh environmental conditions restrict the availability of external goods such as health and wealth, thereby hindering the individual’s capacity for virtuous acts. It follows *ex contrario* that actions that promote goodness in the land simultaneously enhance the accessibility of other preferred indifferents, thereby endowing the land’s inhabitants with a better means to express virtue.

I further argue that these actions are virtuous in themselves. Virtuous action for the Stoics is action that is in accord with nature, and specifically in accord with prudence. It follows that a virtuous person, recognizing the extent to which we benefit from the land’s prosperity, would endeavor to preserve this prosperity. Therefore, actions that promote the good of the land accord with prudence, and thus with virtue.

V. CONCLUSION

The integration of Stoic philosophy with Aldo Leopold’s Land Ethic is a significant development for contemporary environmental ethics. The Stoic concepts of “preferred indifferents” and living “in accordance with nature” are a strong conceptual foundation for the more granular, environmentally oriented ethical framework that Leopold advances, and serve to ground Leopold’s ideas in the Western philosophical tradition. Actions that promote the well-being of the land align with Stoic virtue, and thus realizing these actions becomes a Stoic imperative.

The Stoic and Leopoldian perspectives converge through their mutual recognition of the importance of being a responsible member of the natural order. Although differing in methodology and scope, the

¹⁴ Cicero, *De Finibus Bonorum et Malorum*, 3.16.



mutual consistency of these two schools reflects the enduring nature of the moral tenets that underpin them. The Stoic case for the Land Ethic, as built in this treatment, considers environmental stewardship as an expression of virtue in accordance with nature. In showing this to be the case, I demonstrate that Leopoldian naturalism, often dismissed as too radical or unfounded, has a firm basis in Western thought, which lends it a stronger claim to both academic discourse and real-world practice.¹⁵

¹⁵ Thanks to Alunood Al-Dhawi, Zach Biondi, Joe Coyle, Kevin Crandall, Sophie Kapica, Gabriela Ursu, Helga Varden, and Andrew Schmidt for their input and feedback on prior iterations of this draft.







Jonah Black (he/him) is a senior at the University of Illinois Urbana-Champaign, studying computer science and philosophy. He is broadly interested in virtue ethics, life and death, and topics in science, technology, and society. In his free time, he likes to play and watch soccer, and is an amateur endurance athlete.



RENDERED WITHIN

Epistemological Limits and Radical Humility in Complexity

ABSTRACT

The complexity of the global scale technological totality poses urgent challenges for epistemology and ethics. Drawing from a three-level schema of technology, this paper argues that the level III is characterized by wicked complexity and unpredictable emergent phenomena of which humans are constituent parts, revealing the severe limitations of Enlightenment values and methods. As a consequence, a shift in thinking is necessary. This paper proposes a stance of radical humility, calling for the transformation of individual and systematic ways of thinking to better reflect the complexity of the world.

GAVIN JOHNSON



I. INTRODUCTION: WICKED COMPLEXITY AND THE EPISTEMOLOGICAL DILEMMA

As we are now feeling the full effects of the era of global synchronization through technology, it has become necessary to analyze the full scale of its implications. Issues such as the worsening of climate change, rapid AI integration, and the rise of surveillance are just some of the many technologically related struggles of our age. It is my view that our current methods will not yield meaningful improvements to these phenomena, and that there are many intricacies in the systematic relations of our civilization that make painfully clear our human limitations. I will explain how the technological complexity of our civilization has created a significant epistemological dilemma, one that challenges the Enlightenment values so fundamental to our traditional thinking. To demonstrate this, I will use the analysis by Braden R. Allenby and Daniel Sarewitz in *The Techno-Human Condition* as the foundational ideas for unveiling the epistemological limitations present in complexity.¹ Ultimately, I will argue that these limitations demand a fundamental shift in our ways of thinking that acknowledges epistemological limits as fundamental, holds technological progress as ethically suspect, and dramatically moves common thought away from the values and tendencies of the Enlightenment towards the perspective of *radical humility*.

I will begin with a brief account of the underlying ideas. Allenby and Sarewitz detail three levels of technology characterized by their scale of complexity. The terminology of *technology* is intentionally broad to include the ambiguity of enduring social systems and frameworks, such as governments, languages, systemic racism, methodologies, etc. Allenby and Sarewitz, and many philosophers of technology, reserve the colloquial way of talking about technology (as the plethora of individual physical devices) for other terminology, like *artifact*. Allenby and Sarewitz designate artifacts to the first level of technology, which is described as the “shop-floor level,” where these artifacts are used by individual people, with a clear cause-and-effect chain traceable along intention and consequence.² The second level involves systematic complexity of interconnected and independent artifact-relations into networks, where cause-and-effect relationships begin to become obscured due to the adaptation of the larger system. The third level is the totality of all these technological system-relations together.³

- 1 Braden R. Allenby and Daniel Sarewitz, *The Techno-Human Condition* (MIT Press, 2013).
- 2 Allenby and Sarewitz, *Techno-Human Condition*, 51.
- 3 Allenby and Sarewitz, *Techno-Human Condition*, 37–38, 51, 63.



This hierarchy scheme is admitted by the authors to be “somewhat arbitrary” due to the classification of any technology being dependent on the goals attributed to it.⁴ The gem of this model of thought is the clarity offered by the opposed individual existence of the Level I to the totality of the Level III, as a part-whole relation.

For example, to make sense of this hierarchy of complexity, let us consider a single airplane of Level I. It is a singular artifact that serves a direct purpose of flight and is piloted by an individual who wills it based on linear cause-and-effect relations. Of course, the airplane is made of many subsequent parts and distinct artifacts, but holistically it serves as an understood will-extender in which the pilot does not necessarily need to be aware of the functions of each component to use the whole. Level II would be the broad collection of complex systems which airplanes are a part of or relevant to, such as airline companies, manufacturing industries, trade networks, etc. Level II is a very broad category that best demonstrates the arbitrariness of the hierarchy but is still useful for seeing the increasing intensity of ambiguity that comes along with more complexity. For instance, despite each one of these complex systems related to airplanes of Level II being created for (and by) humans, individuals involved must strategically plan for unexpected issues anticipated to arise. It becomes hard to tell where one thing ends and another begins, how a small change may impact other elements, and how these relations may change over time. This is the loss of linearity, the fuzziness between distinctions of elements and relations, which continues to grow with complexity. This is most apparent at the Level III global scale. This level would be the aggregate of airplane-system-relations, integrated with all other systems. It is the totality in which no single airplane (or individual person) is relevant to the total of the global system.

Level III is the “complex, constantly changing and adapting” collection of systems (both human and natural) which interact in unpredictable ways, producing “emergent behaviors.”⁵ It is characterized by *wicked complexity*, which is described by Allenby and Sarewitz as “when a system’s makeup and dynamics are dominated by differing human values and by deep uncertainty,” and that “any solution to a wicked problem should be expected to create unanticipated but equally difficult new problems.”⁶ This is the concept of *emergence*, in which the complexity of a system yields more than just the sum of its parts; the production of novel phenomena or characteristics.⁷

4 Allenby and Sarewitz, *Techno-Human Condition*, 40.

5 Allenby and Sarewitz, *Techno-Human Condition*, 63.

6 Allenby and Sarewitz, *Techno-Human Condition*, 109.

7 George H., Lewes, *Problems of Life and Mind. First Series: The Foundations of a Creed*. Vol. 2 (James R. Osgood & Co., 1875), 369.



This Level III complexity is wicked in the sense that it is unconquerable by the “favorite child of the Enlightenment, applied reason.”⁸ Any attempt to observe the totality, think up, or apply a solution to one of its many emergent elements, has already failed, as each of these desires has already been facilitated by the complexity of the totality, in which the total form and direction adapts. It is the Ellulian *Technique* with its characteristics of *monism* and the *necessary linking together* of methods under a system-wide sense of rationality.⁹ It is the Deleuzian concept of the *control society*, in which there is a modulated uniformity across all interconnected environments simultaneously. It is a “universal system of deformation,” changing from “one moment to the other” and facilitated by the Level I practical thinking of the technical *one best way*.¹⁰ It is a complete synchronization of systematic relations.

Allenby and Sarewitz detail the nuanced existence of this characteristic in depth. It becomes particularly clear with any analysis of the phenomena mentioned previously: climate change, the internet and social media, and the exponential development and integration of AI technology.¹¹ Complex phenomena like these are both emergent and cultivate unpredictable emergent issues from the complexity of the Level III. There are no direct solutions available at this scale because of the overwhelming relations among many elements.¹² This is the wick- edness of wicked complexity.

This is a situation that must necessarily be addressed in an atypical manner, as current ways of thinking get caught in the individual (Level I) ways of thinking, which serve only to perpetuate the deepening of the totality. How do we go about confronting problems of this scale when our attempts at understanding only aid the necessarily unpredictable complexity of that very problem? If solutions cannot be traced to be simply true-or-false, and we are left with no capacity for trial-and-error, what meaningful answers can our methodologies provide?

II. LIMITATIONS WITHIN THE WEB OF COMPLEXITY

Gilles Deleuze writes about a very similar concept, *control societies*, but with an emphasis on the manipulation of individual behaviors and

8 Allenby and Sarewitz, *Techno-Human Condition*, 111.

9 Jacques Ellul, *The Technological Society*. (Vintage Books, 1964), 79–98.

10 Gilles Deleuze, “Postscript on the Societies of Control,” *October* 59, Winter (1999): 3–7.

11 Susan Schneider, “From LLMs to the Global Brain: The Emergence of Planetary-Scale Artificial Intelligence,” *Disputatio* 16, no. 73 (2024): 87–122.

12 Allenby and Sarewitz, *Techno-Human Condition*, 115, 118.



desires. It is a total system of change that flows within each person, altering desire through the manipulation of all environments simultaneously, and with any opposition being absorbed into it.¹³ This analysis by Deleuze serves to underscore the mechanisms of the Level III totality's influence on the individual who is necessarily within it, showing how it is both a manipulative force and a container of humans upon which it is contingent.

The totality is symbolic of the *oneness* of systems, in contrast to the dissipating individuality of Level I. It is a complete modulation of every individual every element of technological and ecologically integrated systems. It has exponential complexity, where new technological systems are “necessarily linking together” in the mechanisms of interconnected adaptation.¹⁴ Boundaries between perceived distinctions and traceability of cause-and-effect relationships become necessarily fuzzier. The unpredictable emergent phenomena of this scale have urgent issues with real impacts, but we are limited by our relational *position*.

All is contained in the breadth of the umbrella of systematic adaptation and cannot be distinct from it. People are constituents of the totality in the same way that a cell is a constituent of a human body. The cell is not independently relevant to the entire body, but through the coming-together of many cells (of many interconnected system-relations), the totality of the body emerges. The modern, globally synchronized technological world is the birthplace and energy of an organism that we cannot understand because we are individual cells rendered within it.

It is with this relational position in the *web of complexity* that the choice of the individual cannot be meaningfully considered on the scale of the self-deforming totality. An individual choice cannot scale up to any “good” or “bad” result on Level III. The *within-individual* loses relevance as complexity scales, along with the certainty of simplicity.

The common schema for approaching this dilemma is the Enlightenment tendency for more methodological examination, more technological devices, and more techniques of information organization, all under the systematic guise of “progress.” In this sense, the Enlightenment values of rationality, individuality, certainty, and progress are fuel for the fire of growing systematic complexity. These certainly serve well on the practically focused Level I, but it has led to (and further drives us into) increased depth of complexity in which we are *encased* and dependent upon our infrastructure. Through

13 Deleuze, “Postscript on the Societies of Control,” 3-7.

14 Ellul, *Technological Society*, III.



our attempts to tame the complexity of the environment around us, we have unwittingly cultivated wickedness in the systems that have emerged from our interactions as individuals and as institutions. As Martin Heidegger put it, in our pursuit of Truth, the danger remains that the “true will withdraw.”¹⁵

On the scale of Level III, wicked complexity dooms these Enlightenment values to failure. Our very attempts to perceive the totality (much less predictably impact it) cause the total system to adapt its complete form. The result is that the foundations of human understanding, applied directly, are practically useless on this scale. The totality is fundamentally inaccessible to methodology, because our methodologies and institutions are analogous to organs of it.

The recursive relationship of constituent to whole that characterizes wicked complexity may occur outside of human-originated complexity. However, the technological complexity of our current systems is unique in that it cannot be thought of as existing separately for examination. As explained before, the totality emerges from our epistemological, social, political, cultural, economic, and methodological systems interacting on an intensely broad scale with individuals as its constituents. It is an aggregate that emerges from us all as a collective but produces emergent phenomena at the will of no individual. Even so, it is recursively influenced by individuals. It both shapes the context that shapes us and relies on us as constituents to drive the change of total form.

The unnoticed deepening of complexity has likely been facilitated by the survival and habitualization of Enlightenment thought, which, through rationality, holds that all can eventually be understood. As Neil Postman points out, it is this way of thinking that proposes “progress” as the moral good, justified in perceived efficiency.¹⁶ It involves a push toward *objectivity*, rooted in the efficiency of quantification, where any sense of *subjectivity* is pushed away; considered less efficient, less valuable, and less true. This is evidenced in the dominant pursuits of knowledge appealing to static objectivity. However, wicked complexity makes it impossible for any methodology to treat itself as separate from the world it seeks to study. It curates the illusion that knowledge is a universal objectivity in which subjectivity is obscured, and therefore must be discarded, in its efficient pursuit. However, the reality is that no variable can be separated from the entirety.

15 Martin Heidegger *The Question Concerning Technology*, trans. William Lovitt (HarperCollins Publishers), 10.

16 Neil Postman, *Technopoly: The Surrender of Culture to Technology* (Vintage Books), 42, 59–60, 159, 179.



Yet the problems we are facing emerge from the growing complexity of human civilization. How do we address them? Proposed solutions to these problems tend to only involve more delusional technological fixes that add to the complexity. It is this very habitual reliance on the technical problem-solving mentality that justifies the continuation of technological integration, of its progress. This is a wall that our Enlightenment-era thinking cannot penetrate, which only serves to make it more dreadfully looming. This is the epistemological dilemma.

It is in this way that any call for technological progress or for technological solutions should be considered ethically suspect, not for any traceable cause-and-effect relations or intentions, but simply for its blind contribution to the smothering wicked complexity of the totality.

Hans Jonas details that traditional ethics only reckons with “non-cumulative behavior” of the “basic situation between persons.” In contrast, the constant adaptation of the technological totality “constantly overtakes the conditions of its contributing acts,” lacking any linearity in the synchronized manipulation of the Level I context, rendering traditional ethics meaningless.¹⁷ If it is aiding the growing complexity of the totality, it is risking the emergence of uncontrollable, and potentially invisible, characteristics of influence on the individual Level I, and on the ecological systems it integrates with.

This is not to advocate for a complete rejection of technology, but for a more critical lens on what technology is allowed to be integrated, at what intensity, and in what ways. Even so, there does not appear to be any other path other than the technological one. Perhaps it is time to abandon the notion of *paths forward* altogether.

III. RADICAL HUMILITY: NECESSARY THINKING AND NECESSARY PLURALISM

We must begin discussing the next step. It is a challenging step to take. How does a cell, or even a group of cells, protest against the total body without destruction of itself or of the totality? It has been shown that it cannot be known. However, what is known is that we require an altered stance of engagement with the complexity of the world; a novel transformation of our thinking on the Level I scale. It is necessary to shift the tendencies of our individual relations and the framing of the world away from Enlightenment habits of thought. A radical change is

17 Hans Jonas, *The Imperative of Responsibility: In Search of an Ethics for the Technological Age* trans. David Herr (The University of Chicago Press), 7–8.



required in common thinking to consequently alter the whole of the totality in which we together facilitate.

Allenby and Sarewitz work on this difficulty, suggesting some basic principles for navigating the multi-level web of complexity with as much grace as possible. They push for the abandonment of solutions and predictions that are presented under the guise of certainty. They say to focus on constant pluralistic dialogue for frequent decision-making, where choices are of less intensity and of higher frequency.¹⁸ Their underscoring of pluralism being smarter than individuals with expertise has its wisdom in scientific studies of swarm intelligence.¹⁹ It may be principles like these that, despite agitating the nerves of individuality, help us to adjust to the complex web of our relations. Even so, how can this change in thought be meaningfully implemented into the world?

I suspect that any change must have its roots in education (in the broadest sense). However, there is danger in attempting to detail a specific direction for supposed improvement, which implicates some kind of predictive element. Methods of prediction are technological, so attempts at solution only risk aiding the wicked complexity of Level III. This is the danger of the *problem-solving mentality* so prevalent in our thinking, likely woven further by Enlightenment values.²⁰ My concern is that this habit, grouped with the fixation on efficiency and objectivity, filters the potentiality of novel thought outside of technological methods.

Therefore, I am not claiming that there is any bullet list of ideas to be absorbed into our thinking to *solve this problem*. We cannot escape the historical roots of our technological thinking, certainly not through any degree of technological systematization or appeal to rationality. Rather, I will suggest a principle of thinking that continually recognizes the necessary limitations of ideas, methods, ethics, epistemology, etc., in resignation to the reality of the wickedness of the web of system-relations. It is this principle of thinking that encourages a shift in thought away from the current model, of there being a problem to solve, to instead there being *conditions to navigate*. This is the principle of *radical humility*.

This principle is the fundamental rejection of any universal certainty; of it being any more than a technological justification in the totality. It is a continual recognition of ideas, methods, and selves in

18 Allenby and Sarewitz *Techno-Human Condition*, 162–67.

19 Bremer Jörg and Sebastian Lehnhoff "Ant Colony Optimization for Feasible Scheduling of Step-controlled Smart Grid Generation," *Swarm Intelligence* 15, no. 4 (2021): 403–25.

20 Allenby and Sarewitz *Techno-Human Condition*, 31–32.



the relational context of the multi-layered web of complexity, noticing how we as constituents of systems facilitate the emergence of the whole and that all our human elements influence the totality and our consequent perception of it. This is an opposition to the framing of nature as something to be manipulated, as if it were distinct. In a similar way, it is also the opposition to the traditional approach of any method of inquiry being separate from its perceived object of interest.

It is no longer sustainable to traditionally approach any subject of concern as separate; as an objectified distinction accessible by rationality. A perfect, logically sound system of ideas cannot completely account for the unpredictable changes its existence will cause in the context it is rendered within. Any system of ideas that asserts notions of universality, unconditionality, Truth, etc., fails to acknowledge its relevant relational existence in the broader context of the web of complexity. It fails on the principle of radical humility.

This perspective discourages any attempt to detail a specific direction of development because that implies prediction. This is a tricky position, because I am both rejecting *solutions* while also arguing for a principle that is effectively presented as a specific path forward. However, it is not this, nor any one roadmap. It is not radical humility that is the new direction for change. It is not in and of itself a path toward a specific development, but rather a necessary habit of thinking on the individual and institutional levels. It is a change from pursuing false certainty about an environment falsely considered separate into a focus on our own existence, our limitations, and relations within context.

Current thinking and individual relations are the rules (or internal logic of a system) that perpetuate the present existence of the totality. It is analogous to the biological and chemical rules of cells in the human body, interacting to create the whole. If we, the individual constituents, shift our common habits, then the contingent emergent systems must also change in some way. Radical humility is the habituation of doubt to encourage novelty in thinking and wrestle against the systematic forces of total integration.

The current avenues of pursuing knowledge fail when they present any notions of a single *right way*. A siloed-off discipline, no matter the logical strength of its methodologies, cannot grasp the complexity of the entire context. It does not progress toward any Truth, like it is often suggested, but only deepens its justification into institutional systems and common thinking. Often, disciplines are self-contained in their perceived truths and pursuits of knowledge, yielding conflicts when compared to differing approaches. These silos of academia are



increasingly isolated from each other in these passing processes of systematic justifications. What seems to be the methodological pursuit of independent actualizations is just the pursuit of external justification through efficiency. The urgent changes needed to navigate the present cannot be found in the disciplines that have made their continued existence possible as organs of the totality.

If every perspective and approach included the necessary humility due to its recognition of the epistemological limitations of complexity, opportunities would arise for synthesis through difference, rather than disregard. In the rejection of certainty, objectivity, and pure separation through radical humility, we find the possibility of synthesis through accepting perspectives as subjective. Instead of a hierarchy of validity based on systematic power, we acknowledge each system of thought to be equally valid by nature of its logically coherent existence. The pursuit of knowledge then becomes a task, not of finding the *one right way*, but of coherently synthesizing as many perspectives as possible while acknowledging necessary limitations and novelty. The potential is to create a pluralistic map of thought that better resembles the complexity of our interconnected, constantly adapting reality.

IV. CONCLUSION

The complexity-based hierarchy provided by Allenby and Sarewitz provides clarity to the challenges of our technological world. Examined further, it is discovered that we exist as constituents of systems, where the totality of systems reflexively absorbs attempts at understanding into its growing complexity, destabilizing epistemology and ethics. Therefore, we require a philosophical stance of radical humility based on pluralism and the acknowledgment of limitations to better navigate wicked complexity. This stance contrasts with traditional Enlightenment values, but it is not as simple as just picking up a to-do list and incorporating it into our thinking. It is a necessary shift, but one that we cannot aim for without risking furthering the complexity of the totality that needs to be altered. Therefore, work needs to be done to find potential routes of change, likely through a revolutionary analysis into educational processes through the lens of complexity, with methods that account for the principle of radical humility and contextual system-relations.







Gavin Johnson (he/they/she) is a senior at Georgia Southern University Armstrong, studying philosophy, English, and mathematics. Their main interests are in education, process philosophy, complexity science, and queer-feminist theory. Their goal is to become an educator, ideally through pursuing their PhD in philosophy.



NISHIDA KITARŌ'S PURE EXPERIENCE BEYOND CONSCIOUSNESS

ABSTRACT

This paper examines Nishida Kitarō's concept of pure experience and whether one can be conscious of it. Pure experience being the foundation for all experience in that it merges the subject and object while rejecting common-sense dualism. However, while tracking Nishida's argument, it becomes apparent that there is a contradiction as he characterizes thinking as "deliberate discrimination" and "a kind of pure experience." I trace Nishida's movement from a psychological to a metaphysical account of pure experience. I argue that one cannot be conscious of it without reintroducing the subject-object distinction it precedes.

AMY ESPARZA



I. INTRODUCTION

Western epistemology has been historically governed by a dualistic model that divides the individual and reality into two distinct camps: realism and idealism. Scholars, such as Kopf, argue that Nishida Kitarō believed that this dualism permeated all philosophical discourse — even monism, which defines itself against dualism and falls prey to binary thinking.¹ This led Nishida to construct an answer that would challenge the framework altogether. As Abe notes in his introduction to *An Inquiry into the Good*, “[Nishida] had to transform Zen experience into a philosophical answer,” and thus Nishida introduces the concept of pure experience, a metaphysical phenomenon that inverts all our intuitive conceptions of the world.² It is a concept defined by two features: it merges the subject and object while rejecting common-sense dualism, and it is pre-cognitive, occurring prior to reflection and conceptualization.³ This is counterintuitive, as it suggests that pure experience is non-conscious yet is meant to replace our usual notion of consciousness. A way to make this more comprehensible is found in Daisetz T. Suzuki’s “Zen and Swordsmanship I,” where he explains how, “when the opponent tries to strike you, your eyes at once catch the movement of his sword, and you may strive to follow it. But as soon as this takes place, you cease to be master of yourself, and you are sure to be beaten. This is called ‘stopping.’”⁴ The swordsman is in a flow state, acting in the moment without reflection. He is purely instinctual, exemplifying pure experience in action. The difficulty of pure experience is that it is a concept that undermines the tools you would normally use to grasp it and asks for an understanding that cannot exactly be put into language.

Yet as one dives deeper, apparent inconsistencies within the framework emerge that one must grapple with to see whether this is an adequate conceptualization of what our understanding should be. For example, Nishida characterized thinking as “deliberate discrimination”⁵ that destroys the concrete unity of pure experience but later

1 Gereon Kopf, “Between Identity and Difference: Three Ways of Reading Nishida’s Non-Dualism,” *Japanese Journal of Religious Studies* 31, no. 1 (2004): 73–103, <https://www.jstor.org/stable/30233738>.

2 Masao Abe, “Introduction,” in *Inquiry into the Good*, by Nishida Kitarō, trans. Masao Abe and Christopher Ives (Yale University Press, 1990), vii–xxvi.

3 Kitarō Nishida, *An Inquiry into the Good*, trans. Masao Abe and Christopher Ives (Yale University Press, 1990), 47.

4 Daisetz T. Suzuki, “Zen and Swordsmanship I,” in *Zen and Japanese Culture* (Princeton University Press, 1959), 87–136, <https://www.jstor.org/stable/j.ctvc77449.12>.

5 Nishida, *Inquiry into the Good*, 3.



says that “the activity of thinking constitutes a kind of pure experience.”⁶ This leads into the question: Can we become conscious of pure experience? Alternatively, is this question incoherent with Nishida’s framework? In this paper, I will argue that Nishida’s *Inquiry into the Good* systematically inverts the assumption that consciousness is the primary function for pure experience, and that we can never truly be ‘conscious’ of pure experience. The name itself could reasonably leave one with the impression that this is a type of experience one can perceive or reflect upon, yet through the examination of pure experience, it will be made clear that it is truly the fundamental field of reality that places the individual’s consciousness as a secondary component within it. The self is not the “owner” of experience, but a temporary focal point within the event of experiencing itself. In doing so, Nishida moves from constraining pure experience to the individual, to shifting pure experience as the medium through which reality takes place, creating a relational unity between the two. As a result, it is impossible to be conscious of pure experience, and framing the concept in such a way would be a misunderstanding of the system in the first place. It is important to note that Nishida’s conception of pure experience arises in his earliest work, meaning we are dealing with what is often called ‘early Nishida.’ This stage is characterized by his attempt to articulate a philosophical foundation for insights drawn from Zen practice without explicitly invoking Buddhist terminology.

II. THE INDIVIDUAL’S RELATION TO PURE EXPERIENCE

To understand whether an individual can be conscious of pure experience, there must first be an understanding of where the individual stands with relation to it. In working toward the final definition of pure experience, Nishida characterizes it as having “no meaning whatsoever; it is simply a present consciousness of facts just as they are.”⁷ This type of experience is meant to be the simplest experience one can have and is foundational for all other experiences. It occurs prior to the subject-object distinction, when the “experiencer” has not yet been able to separate themselves from what they are experiencing; rather, they are still one and the same. As a result, Nishida expands our understanding by introducing the “perceptual train,” describing how a person can move through actions without breaking the subject-object distinction; they act without reflection. This shows that this phenomenon is not restricted to simple sensations; rather, the defining

6 Nishida, *Inquiry into the Good*, 13.

7 Nishida, *Inquiry into the Good*, 4.



aspect is that pure experience is a strict unity of concrete consciousness. A unity that leaves no room between “the demands of the will and their fulfillment.”⁸ Pure experience is fundamentally connected to the individual; however, at this point, it is unclear whether that individual could be conscious of it.

At this point, Nishida provides what seems like an answer: he says that “the activity of thinking constitutes a kind of pure experience.”⁹ A statement that, until this point, must be incompatible with what he has already worked so hard to establish. As early as the first page, thought was characterized as a “deliberate discrimination” within experience, and that characterization primed our minds to understand the nature of this concept.¹⁰ Nishida does not appear to be trying to contradict that understanding but instead is now expanding the scope of what pure experience can be. Still, we now need to reconcile how both statements can be true. How can thinking be both a form of discrimination and a kind of pure experience?

III. THE PSYCHOLOGICAL ANSWER

Initially, we are getting what I propose to be a ‘psychological answer.’ I claim this because it focuses solely on the individual’s mindset and lays the foundation for what will later become a ‘metaphysical shift.’ This distinction is necessary because it constitutes only part of pure experience and explains why the phenomenon needed to be expanded in the first place. In this context, thinking means “the activity that determines relations between representations and unifies them.”¹¹ To think of thinking as part of the pure experience already established, he highlights how thinking can be passive, much like the passivity we experience with our perceptions. This psychological account, however, proves insufficient because it relies solely on the individual, thereby ignoring the constitutive role of reality itself in the phenomenon of pure experience.

In this account, pure experience is an individual having experiences, completely constrained to themselves, effectively missing the playing field on which these experiences take place. It makes it seem as though pure experience is completely individualistic, disregarding the fact that the reality we interact with is a pivotal component of this phenomenon. The psychological reading presupposes the very sub-

8 Nishida, *Inquiry into the Good*, 8.

9 Nishida, *Inquiry into the Good*, 13.

10 Nishida, *Inquiry into the Good*, 3.

11 Nishida, *Inquiry into the Good*, 11.



ject-object distinction that pure experience is meant to precede. It assumes that there is a subject that has experiences, and those experiences are objects of analysis or reflection. Even if one finds themselves experiencing something that was pre-reflective phenomenologically, that experience is still post-reflective conceptually. Therefore, the psychological account cannot be the ground of pure experience; instead, pure experience is the condition for the possibility of psychology as such. The limitations of the psychological account, therefore, necessitate a metaphysical shift in Nishida's conceptualization, as we still do not have an answer that unites the fact that thinking can be discriminatory while also being part of pure experience without creating a contradiction.

IV. THE METAPHYSICAL SHIFT

My analysis will now show how Nishida responds to this tension not psychologically, but metaphysically. As Abe notes, Nishida “regarded pure experience as the sole reality and wanted to develop his philosophy on this basis; he was not satisfied with the theories of pure experience expounded by the psychological philosophers,” and therefore sought to “connect his own theory of pure experience to transcendent philosophy or metaphysics by grasping experience as active.”¹² In other words, the metaphysical shift is not an interpretive imposition, but a direction Nishida himself considered necessary. By reacting to these “psychological philosophers,” there is concrete signaling that, while informed by these philosophies, the next direction Nishida's answer must take will separate itself from the commonsense dualism that inherently underlies any account relying on the individual.

Nishida's metaphysical shift does not so much resolve the problem of pure experience as it relocates it. This relocation becomes especially clear when his position is set against the psychological framework of William James. Nishida himself acknowledges that the initial, more psychological move of identifying thinking within pure experience came from William James. Yet the contrast between their final positions is striking. In “The World of Pure Experience,” James writes, “There is no other nature, no other whatness than this absence of break and this sense of continuity in that most intimate of all conjunctive relations, the passing of one experience into another when

¹² Abe, “Introduction,” xiv.



they belong to the same self."¹³ While James affirms the continuity of experience, he restricts its significance to what is immediately given; it "rests on the claim that experience is not a thing but a *function*."¹⁴ Whereas Nishida takes pure experience to be the ground of a reality that surpasses the individual. This divergence specifically reveals how Nishida situates aspects of his own Zen philosophy into this broader philosophical conversation, transforming a psychological notion into a metaphysical principle. Both approaches contrast with commonsense dualism, but whereas James focuses on the pure experience of being a unity within the individual, Nishida makes the unity something the individual exists within. Revealing how the problem of pure experience should not have been focused within the individual; rather, it should go beyond the individual.

From Nishida's perspective, pure experience becomes the basis for experience, and he says, "It is not that there is experience because there is an individual, but that there is an individual because there is experience. The individual's experience is simply a small, distinctive sphere of limited experience within true experience."¹⁵ This inverts our traditional assumption that pure experience is not a psychological state produced by an individual brain; it is fundamental to an objective reality from which the world and individual consciousness arise. It prevents the argument from devolving into idealism and instead fosters pure experience and thinking, which are interdependent. In this way, thinking stops being an activity *opposed* to pure experience and becomes part of it, and the act of discrimination occurs within the unity of experience rather than functioning as a disruption.

To avoid interpreting this view as eliminating the individual, Nishida clarifies the person's role within true experience. Pure experience is something that contains individual people; however, it does not mean that the objective reality is pure experience; the individual still plays a pivotal role in this phenomenon. 'True experience' cannot merely be the world in and of itself; the mind's interaction with it is still a key aspect of 'true experience.' Instead, the individual plays the role that makes the field determined and self-aware. Therefore, the mind's interaction with the world is not a secondary event but is the primary substance of reality itself; 'true experience' is pure experience in this

13 William James, "A World of Pure Experience," *The Journal of Philosophy, Psychology and Scientific Methods*, vol. 1, no. 20 (1904): 533–43, <https://www.jstor.org/stable/2011912>.

14 Andrew Feenberg and Yoko Arisaka, "Experiential Ontology: The Origins of the Nishida Philosophy in the Doctrine of Pure Experience," *International Philosophical Quarterly* 30, no. 2 (June 1990): 180, <https://doi.org/10.5840/ipq19903023>.

15 Nishida, *Inquiry into the Good*, 19.



final form. My understanding is that true experience is the non-dual event of interaction, prior to the conceptual split into mind and world.

This resolution invites a pressing objection, if ‘true experience’ is fundamental and we are merely ‘limited spheres’ within it, does Nishida account for individuality and a subjective perspective? No, for Nishida, unity is not sameness, but structured relatedness within a shared system. This is clear when he discusses the systemic nature of consciousness, saying, “If we acknowledge that my consciousness of yesterday and today are independent and at the same time one consciousness in that they both belong to the same system, then we can recognize the same relationship between one’s own consciousness and that of others.”¹⁶ It establishes that, for Nishida, unity is not always *sameness*; rather, it is systemic relatedness. Nishida exemplifies this by asking us to consider an influential person who can sway many people’s opinions. This does not necessarily mean everyone will be swayed in the same way or to the same degree; however, it is possible to be swayed because we all exist in the same system. It is possible for a great person to influence everyone, and for a mathematical principle to hold true every time.¹⁷ Rather, the system’s unity comes from its structure, which allows us to hold diverse opinions while working within it. If we were not working within the same structured system, it would be impossible for a great person to influence anyone, and, by extension, for us to communicate with one another.

V. THE ESSENCE OF CONSCIOUSNESS

This clarified account of individuality is essential for addressing whether we can be conscious of pure experience at all. If pure experience is the non-dual field in which both the objective world and the individual arise, then it cannot appear as an *object* to a subject. A common concern is that the metaphysical account dissolves individuality altogether, but Nishida’s systemic conception of unity shows that individuality is preserved as a standpoint within a larger relational whole. This unity aligns with our everyday experience of acting in a world that is neither strictly subjective nor strictly objective.

With this in mind, we can now ask whether the individual, understood as a limited sphere within true experience, can reflect upon pure experience itself. Masao Abe’s commentary makes the answer clear. Abe states that “the standpoint of consciousness [is] truly the subject,

¹⁶ Nishida, *Inquiry into the Good*, 44.

¹⁷ Nishida, *Inquiry into the Good*, 45.



not the object, of consciousness," and therefore, you cannot make consciousness into an object of consciousness without already presupposing consciousness.¹⁸ Because pure experience is the pre-reflective ground of consciousness, attempting to become *conscious* of it reintroduces the very subject-object division; it precedes. Thinking back to the example of the samurai, they are aware and conscious in the sense that they are not asleep, but they are not conscious in the sense that they can reflect on what is happening to them without breaking the unity of consciousness. Thus, the question of being conscious of pure experience is based on a misunderstanding of its fundamental nature: pure experience is not something we can observe, but the condition that makes all observation possible.

VI. THE OBJECTION: CAN WE BE CONSCIOUS OF PURE EXPERIENCE

Yet the objection still stands: if the act of discrimination occurs within the unity of experience rather than disrupting it, then why would becoming conscious of that unity be any different? In fact, when discussing the lack of distinction between consciousness and reality, Nishida himself says, "True reality is not found outside of phenomena of consciousness, and the true reality realized in direct experience always comes forth in the same mode."¹⁹ For there to be a reality, there must be consciousness; the two are not independent of one another. Thus, to say one cannot be conscious of pure experience, the domain for which consciousness arises, seems contradictory to Nishida's conception of reality itself. In fact, Nishida even says in the very beginning of his work on pure experience that, "when one directly experiences one's own state of consciousness, there is not yet a subject or an object, and knowing and its object are completely unified"²⁰ Making it clear that we can be conscious of pure experience.

For instance, we can experience the state of consciousness, and if this consciousness occurs in the unity required for pure experience, then we can be conscious of pure experience. From this, one might conclude that an individual can be aware of pure experience. If we accept that there is no distinction between the individual and reality, then perhaps our mere existence is consciousness of pure experi-

18 Masao Abe, "The Logic of Absolute Nothingness as Expounded by Nishida Kitarō," *The Eastern Buddhist* 28, no. 2 (1995): 168, <https://www.jstor.org/stable/44362093>.

19 Nishida, *Inquiry into the Good*, 54.

20 Nishida, *Inquiry into the Good*, 3-4.



ence. When you take this metaphysics and carry it to its conclusion, we must accept that we can be conscious of pure experience, since to say otherwise presupposes there is an individual that undermines the goal of unity in the first place.

However, I argue that concluding consciousness grasps pure experience would be to mistake a part for the whole. Consciousness arises from pure experience, but that does not grant it access to the totality that makes it possible. On the other hand, looking back at our definition, pure experience merges the subject and object while rejecting common sense dualism, and it is pre-cognitive, occurring prior to reflection and conceptualization.²¹ Under this definition, even an understanding of consciousness based on awareness rather than reflection is incompatible with pure experience and reinforces my point that one cannot be conscious of pure experience. Granting that awareness is consciousness of pure experience would risk devolving back into common sense dualism, which Nishida was avoiding in this conception in the first place.

In “Experiential Ontology: The Origins of the Nishida Philosophy in the Doctrine of Pure Experience,” this interpretation is supported. Here, Feenberg and Arisaka state, “The subject does not ‘have’ the pure experience ‘of’ the object, but rather pure experience is a unified totality, and subject and object come to be differentiated within that totality through reflection.”²² Consciousness exists within pure experience, whereas the act of being conscious requires a reflective state that breaks concrete unity and eliminates the state of pure experience. So, while it may be true that we can be conscious and exist within pure experience, that does not mean we are conscious of pure experience.

VII. CONCLUSION

To conclude, pure experience is not something we can be conscious of; it is a unity between ourselves and constitutes the way we are able to experience anything in the first place. While thinking, constituting a type of pure experience may lead some to believe that this means we can reflect upon pure experience and therefore be conscious of it. This would be a misunderstanding of what Nishida means by thinking and does not account for the phenomenon in its entirety. By tracing his movement from a psychological answer to a metaphysical account in which the individual becomes a “limited sphere” within true experience,

²¹ Nishida, *Inquiry into the Good*, 47.

²² Feenberg and Arisaka, “Experiential Ontology,” 178.



rience, we have seen why pure experience cannot be an object of consciousness. Yet this shift raises further questions: Was Nishida right to make this move? Does this early metaphysical expansion adequately secure non-duality, or does it risk obscuring the immediacy he first set out to describe? I propose that the metaphysical shift was necessary. As outlined in this paper, the mere psychological account introduced tension with his original emphasis on immediacy. However, that tension is not directly resolved in his early work, though the metaphysical conception of pure experience sets the foundation for an understanding that avoids dualism altogether. Thus, these questions are precisely what lead us into Nishida's later philosophy, such as his logic of basho, his reflections on absolute nothingness, and his evolving account of self-awareness. All of these work to rectify the inconsistencies in this earlier work and solidify the fact that the tensions addressed in this work are tensions Nishida identified as well.







Amy Esparza (She/Her/Hers) is a 4th year student at California Polytechnic State University- San Luis Obispo. She is pursuing a degree in Philosophy with a Law and Society minor. Her philosophical interests include Phenomenology, Latin American philosophy, and Feminist ethics. In her free time, she enjoys bowling. Next year, she will be an elementary school teacher.



WHY CONSCIOUSNESS DOESN'T NEED COMBINING

ABSTRACT

This paper argues that Philosophy of Mind's Combination Problem fails to explain how many micro-experiences can form a unified subject. Instead, the paper proposes Suhrawardī's illuminationist metaphysics as a solution, which consists of a graded monism in which reality is one substance—Light—that just varies in intensity. This reframes consciousness as intensification rather than combination. The conclusion holds that unity is fundamental, not constructed, and that this framework offers a simpler, more coherent alternative to combinatorial panpsychist models.

AYESHA KAYANI



I. INTRODUCTION

Panpsychism is the philosophical view that consciousness is a fundamental feature of reality. According to this view, even the smallest components of the world possess extremely simple forms of experience. It continues to draw interest in the philosophy of mind because it avoids the sharp explanatory jump required by standard physicalism. Physicalism is the dominant view in contemporary philosophy of mind, which holds that everything that exists is ultimately physical. On this account, consciousness arises from physical processes in complex systems such as the brain.

Rather than claiming that consciousness somehow appears when matter becomes complex enough, panpsychism holds that very simple forms of experience are built into the fundamental constituents of the world. Yet the view faces a central obstacle: the Combination Problem. The Combination Problem refers to the challenge of explaining how many simple forms of consciousness could combine to produce the unified and complex consciousness experienced by human beings. Even if the smallest parts of reality possess minimal forms of experience, it remains unclear how these countless micro-experiences could cause the rich and unified consciousness characteristic of human life.

This paper argues that a powerful way around these difficulties emerges from an unexpected source: the illuminationist metaphysics of the twelfth-century Persian philosopher Shihāb al-Dīn Suhrawardī. Suhrawardī does not attempt to solve the Combination Problem by supplying a better mechanism for fusing micro-subjects. Instead, he challenges the starting assumption that reality is fundamentally many. His illuminationist metaphysics is built around a single, graded substance—Light (nūr)—which he describes as intrinsically self-revealing.¹ Everything that exists is a manifestation of this one luminous field, differing only in degree of intensity. Here, the idea that consciousness must be combined is already a misunderstanding; unity is built into the structure of reality from the outset.

The aim of this paper is not to offer a historical interpretation of Suhrawardī, but to draw on his metaphysics as a conceptual resource for contemporary debates about consciousness. The discussion begins by examining the structure of the Combination Problem and explaining why existing panpsychist approaches struggle to account for sub-

1 Muhammad Umar Riaz Abbasi, "Philosophy of Illuminism: Comparative Study of Shahab al-Din Suhrawardi and Mulla Sadra," *Migration Letters* 21, no. S10 (2024), 4.



jecthood and phenomenal unity. It then turns to Suhrawardī's account of graded Light, showing how his monistic ontology replaces the assumption of many independent subjects with a single, unified reality. The paper next explains how the distinction between dependent and self-subsisting degrees of Light yields an account of subjecthood that avoids any appeal to aggregation. Finally, it addresses a series of objections and argues that Suhrawardī's position can respond to each in a coherent and principled way.

By reframing the entire picture of reality rather than working within it, Suhrawardī offers a way around the Combination Problem. The puzzle disappears once we surrender the idea that consciousness must be built from independent parts.

II. DIAGNOSING THE COMBINATION PROBLEM

A clear understanding of the Combination Problem is essential for seeing why current versions of panpsychism struggle and why a different starting point may be required. The problem is not a single puzzle but a cluster of connected difficulties about how subjects and experiences could ever fit together. No matter how we picture the relationship between the simplest forms of consciousness and the complex consciousness of a human being—whether as a literal sum of parts or as a higher-level pattern—the same binding problem remains.

We can break the Combination Problem into three main challenges:

1. THE SUBJECT-SUMMING PROBLEM

If the basic units of the world are subjects of experience, then anything made of many such units—from a rock to a brain—would contain many subjects. The human brain, for instance, is built from trillions of particles. If each of those particles is a subject, why does a person experience themselves as exactly one subject rather than a vast crowd? Simply arranging many subjects together does not explain how one unified subject could emerge.

2. THE PHENOMENAL UNITY PROBLEM

Even if we set subjects aside and focus only on experiences, difficulty remains. Human consciousness feels like a single, seamless field. We see, hear, and think within one interconnected awareness. But if micro-experiences are tiny, self-contained points of experience belonging to different micro-entities, why should they ever form a unified whole? The situation is like trying to imagine how isolated



pixels, each with no knowledge of the others, could somehow produce a coherent image.

3. THE GRAIN AND PALETTE PROBLEM

Finally, there is the issue of qualitative richness. The experiences attributed to fundamental entities are supposed to be extremely simple, perhaps nothing more than vague, primitive sensations. How could combining a large number of such simple experiences produce complex ones, like the taste of wine or the feeling of nostalgia? Adding simple things together does not automatically generate something more sophisticated.

A common response to these challenges is to propose a special bonding relation, which is meant to fuse the micro-level experiences or subjects into a single unified one.² However, this has problems of its own. The relation is usually presented as a primitive feature of reality, introduced solely to do work that the theory cannot explain on its own. This adds additional complexity to the ontology—now we have micro-subjects and an unexplained bonding mechanism—and still leaves us without a clear understanding of how the combination really works.

Together, these difficulties suggest that the Combination Problem may not arise from consciousness itself but from a particular metaphysical picture that assumes reality is fundamentally made of many subjects. Suhrawardī's illuminationist framework challenges this assumption directly. By rejecting this inherited picture at its root, his ontology offers a way to dissolve the Combination Problem rather than attempting to solve it piecemeal.

III. SUHRAWARDĪ'S ONTOLOGICAL BYPASS

Shihib al-Dīn Suhrawardī (1154–1191), the founder of the Illuminationist (Ishrāqī) school of philosophy, developed his metaphysics within the rich intellectual environment of the medieval Islamic world. Earlier philosophers in this tradition, particularly those influenced by Aristotle, had emphasized logical reasoning and abstract metaphysical analysis.³ Suhrawardī sought to complement this approach by grounding philosophy in the concept of Light (nūr), which he un-

2 Philip Goff, *Consciousness and Fundamental Reality* (Oxford University Press, 2016), 8.

3 Peter Adamson, *Philosophy in the Islamic World: A Very Short Introduction* (Oxford University Press, 2015), 33.



derstood as the fundamental principle of reality.⁴ For him, Light represents pure presence—to exist is to be present or luminous in some degree. Although his philosophy has implications for many areas, including epistemology and cosmology, this paper focuses specifically on how his account of graded Light can contribute to contemporary debates about consciousness.

Rather than solving the Combination Problem on its own terms, Suhrawardī's philosophy offers an ontological bypass that dissolves its foundational premises. His framework does not explain how many subjects combine because it denies that reality is, at its base, a plurality. Suhrawardī saw himself as a reviver of a perennial wisdom who combined the rationalistic philosophy of the Peripatetics, the practical wisdom of the Sufis, and the intellectual intuition of the Ishrāqī tradition.⁵ It is from this ambitious synthesis that a new ontological starting point emerges.

The core of Suhrawardī's metaphysics is that all existence is a single, self-evident substance he calls Light (nūr).⁶ This is not physical light but a metaphysical concept for pure, self-aware presence. For Suhrawardī, to be real is to be luminous, and to be luminous is to be present to oneself. This foundational substance—this principle of self-evidentiality—is precisely what contemporary panpsychists are seeking, which is a non-physical experiential ground of reality.

Monism is the view that reality is ultimately composed of a single underlying substance or principle rather than many independent entities. From this monistic premise, Suhrawardī introduces his doctrine of *tashkīk al-nūr*, or the gradation of Light.⁷ In this model, differences between beings are not differences of substance but of *intensity* or *degree of luminosity*. A rock, a plant, and a human are not different kinds of things but are all manifestations of the one Light, distinguished only by the intensity of their reality. A brighter light and a dimmer light are not two entities, but the same reality at different scalar values. This doctrine of graded light proved so powerful that later thinkers, such as the 15th-century philosopher Ibn Turka, would extend it into a comprehensive “lettrist metaphysics” to explain the structure of language itself as a manifestation of divine luminosity.⁸

4 Shihāb al-Dīn al-Suhrawardī, *Ḥikmat al-Ishrāq* (Dār al-Ma‘ārif al-Ḥikmah, 2010), 9.

5 John Walbridge, *The Wisdom of the Mystic East: Suhrawardī and Platonic Orientalism* (State University of New York Press, 1992), 121.

6 Al-Suhrawardī, *Ḥikmat al-Ishrāq*, 24.

7 Al-Suhrawardī, *Ḥikmat al-Ishrāq*, 90.

8 Leonard Lewisohn, “Ibn Turka and the Illuminationist Roots of Hurufi Lettrism,” in *The Heritage of Sufism*, vol. 3, ed. Leonard Lewisohn and David Morgan (Oneworld, 1999), 23–45.



This framework reassesses the discussion entirely. The Combination Problem begins by assuming that reality is made up of many separate micro-subjects whose experiences must somehow be brought together. Suhrawardī's graded monism rejects this assumption from the outset. In his view, there is not a collection of independent subjects but a single luminous reality that appears in different degrees. The central question, therefore, shifts from "How do many subjects combine into one?" to "How does one underlying field of Light present itself in varied intensities?" Once the premise of fundamental plurality is abandoned, the demand for a special binding mechanism no longer arises. This reframing has important consequences for how we understand subjecthood. The task is not to explain how individual units are fused, but how different levels of intensity within one continuous reality give rise to different forms of awareness.

IV. REFRAMING SUBJECTHOOD AND INTRINSIC UNITY

Within a monism of Light, a "subject" cannot be something possessed by separate, fundamental units of reality. Subjecthood must instead be understood dynamically, as a function of the varying degrees of luminosity that make up all existence. Suhrawardī's system gives us a clear way to frame this, and he distinguishes between two kinds of manifestation:

SELF-SUBSISTING LIGHTS (NŪR JAWHARĪ)

These are points within the luminous field where Light reaches a high enough intensity to become self-aware. A subject is not a basic building block of reality, but it is the result of Light achieving a certain threshold of clarity and presence. Subjecthood is an accomplishment of intensity, not a property of independent parts.⁹

DEPENDENT LIGHTS (NŪR 'ARADĪ)

These are lower degrees of luminosity. Still forms of phenomenal presence, but not strong enough to sustain self-awareness on their own. They appear, but they do not appear to themselves. They are experiential in nature but not subjects.¹⁰

This distinction allows Suhrawardī's framework to avoid the Subject-Summing Problem altogether. If only high-intensity lights are true subjects, the world is not populated by trillions of micro-subjects that must somehow merge. Subjecthood emerges only where

9 Shihāb al-Dīn Suhrawardī, *The Philosophy of Illumination*, trans. John Walbridge and Hossein Ziai (Brigham Young University Press, 1999), 79.

10 Suhrawardī, *Philosophy of Illumination*, 45.



the one luminous substance reaches the appropriate intensity. Below that point, there is experience, but there are no separate subjects to combine.

The same structure also dissolves the Phenomenal Unity Problem. This problem assumes that experience begins as many discrete pieces, which must then be stitched into a unified whole. For Suhrawardī, all phenomenal reality is already a modulation of a single, continuous field of Light. Unity is not something that needs to be achieved, as it is the default condition. What look like scattered micro-experiences are simply lower-intensity expressions of the same underlying presence. The need for a binding mechanism disappears because the experiences were never metaphysically separate in the first place.

V. CONSCIOUSNESS WITHOUT ASSEMBLY

The Combination Problem rests on a metaphysics of assembly, where a conscious subject is understood as something built by adding together many smaller conscious parts. In this view, macro-consciousness is the result of quantitative aggregation. Suhrawardī's system points toward a fundamentally different model: one that treats consciousness not as something assembled, but something realized through qualitative intensification.

The contrast can be made clear with a simple analogy. Screwing additional lightbulbs into a room increases brightness by adding more separate sources—this is quantitative aggregation. Turning up a dimmer switch, by contrast, increases brightness by intensifying a single source. Suhrawardī's model follows the latter. Macro-consciousness is not a room filled with countless tiny lights, but a single light brought to a high degree of clarity and determination.

In this view, human consciousness is not the sum of billions of micro-consciousnesses. It is the single substance of Light manifesting at a locally maximal degree of intensity within a suitably organized system, such as the human brain. The brain does not combine many independent subjects; rather, its structure allows the underlying phenomenal reality to appear with greater determinacy, yielding a complex field of experience.

Understanding consciousness in terms of qualitative intensification dissolves the remaining combination worries without appealing to mysterious bonding relations. The Unity Problem disappears because unity is already present in the single luminous substance



itself, rather than something that must be constructed. Likewise, the Grain and Palette Problem is resolved because experiential richness is not built from simple components but rather increases with intensity—a stronger manifestation of Light is, by its nature, a more fully articulated one.

Although this Suhrawardian picture is elegant and parsimonious, its adequacy cannot simply be assumed. The next step is to test it against a series of objections that challenge whether intensification can really replace combination as an account of subjecthood and phenomenal structure.

VI. OBJECTIONS AND ILLUMINATIONIST REPLIES

This section serves as a crucial test of the coherence of the Suhrawardian model. A framework that dissolves one problem only to create others is no real advance. Here, we address three potent objections and provide replies grounded in the illuminationist framework.

OBJECTION 1

One might object that Suhrawardī merely relocates the problem of emergence rather than eliminating it. The transition from a non-subjective dependent light to a self-subsisting light that qualifies as a subject can appear to involve a sharp and unexplained leap. If dependent lights lack subjecthood altogether, then the appearance of a genuine subject seems to reintroduce the very kind of brute emergence that panpsychism seeks to avoid.

However, this objection rests on a categorical misreading of Suhrawardī's ontology. Dependent and self-subsisting lights are not different kinds of entities. They are different degrees of the same luminous substance. Subjecthood is not a new property but emerges when this single substance reaches a sufficient level of intensity and self-presence. The transition is therefore continuous rather than abrupt, avoiding any jump from the non-experiential to the experiential.

OBJECTION 2

A second concern is that the idea of intensification itself may appear as though an unexplained rule of the theory. If subjecthood appears whenever Light reaches a certain level of intensity, this might seem no better than the panpsychist's move of simply postulating a



special “bonding relation” to connect micro-experiences.¹¹ In both cases, critics might say, an unexplained principle is doing important work.

However, the two views are not on equal footing. Combinatorial panpsychism starts with many separate micro-subjects and then adds an extra relation to bind them together. This increases the number of basic assumptions the theory must make. Suhrawardī’s framework, by contrast, begins with just one underlying reality and one internal principle that explains how it varies. The same principle that accounts for different degrees of existence also accounts for different degrees of awareness. Even if intensification is taken as basic, it offers a more viable explanation because it avoids adding a separate mechanism solely to solve the combination problem.

OBJECTION 3

A final objection concerns the status of the physical world. If reality is fundamentally a single substance of Light, it may be unclear how solid, spatially structured objects arise at all. Pure phenomenal presence seems insufficient to ground the features studied by physics such as spatial extension, dimensional structure, resistance, and causal interaction.¹²

However, Suhrawardī’s ontology includes an essential dimension. Alongside Light, Suhrawardī recognizes *barzakh*, often translated as “dusky barriers.”¹³ These are not a second substance added to Light, rather, they refer to regions where the expression of Light is limited or dimmed. In Suhrawardī’s metaphysics, *barzakh* functions as a boundary that constrains the expression of Light. When luminosity is limited, it gives rise to the features we associate with the physical world, such as spatial extension, resistance, and material structure. Matter, therefore, is not a separate substance opposed to consciousness but arises where Light is partially dimmed.

This concept also plays a broader role within Suhrawardī’s philosophy. By explaining physical reality as a limitation of luminosity rather than a separate material domain, *barzakh* helps maintain the unity of his metaphysical system. All levels of reality, from pure Light to physical bodies, remain part of the same underlying ontological framework, differing only in the degree to which luminosity is present or constrained.

11 Philip Goff, “Why Panpsychism Doesn’t Help Us Explain Consciousness,” *Dialectica* 63, no. 3 (2009): 289–311, 291.

12 James Ladyman and Don Ross, *Every Thing Must Go: Metaphysics Naturalized* (Oxford University Press, 2007), 19.

13 Al-Suhrawardī, *Ḥikmat al-Ishrāq*, 9.



VII. CONCLUSION

The Combination Problem stands as arguably the most significant obstacle to the acceptance of panpsychism. This paper has argued that the 12th-century illuminationist metaphysics of Shihāb al-Dīn Suhrawardī offers a powerful framework for bypassing this problem entirely. By replacing the premise of ontological plurality with a graded monism of a single phenomenal substance—Light (*al-nūr*)—Suhrawardī's system dissolves the very questions upon which the problem is built.

The main strength of this model is its simplicity. It explains subjecthood and the unity of experience without adding special bonding relations by starting from a single, unified reality. In this framework, subjecthood arises through degrees of intensity, and unity is the default rather than something that must be built. Seen this way, problems like subject-summing, lack of unity, and limited experiential richness are not deep puzzles about consciousness itself, but consequences of starting from a more complicated metaphysical picture.

Adopting this view does require strong metaphysical commitment. Accepting a form of idealist monism in which phenomenal presence is fundamental. This is not a flaw of the position, but a trade-off. In return, it offers a simpler and more coherent alternative to the ongoing difficulties faced by combinatorial versions of panpsychism. If this illuminationist framework proves promising, several avenues for further inquiry remain open. One important question concerns how the idea of graded luminosity might relate to the organization of complex biological systems such as the brain. While this paper has argued that consciousness should be understood as an intensification of a single luminous reality rather than a combination of many subjects, further work would be needed to clarify how particular structures enable higher degrees of such intensity. More broadly, engaging with Suhrawardī's philosophy suggests that historical and non-Western metaphysical traditions may contain concepts that can contribute greatly to contemporary debates about the nature of mind and reality.







Ayesha Kayani (she/her) is a philosophy student graduating in 2026 from the University of Liverpool. She is interested in Philosophy of Mind, Quantum Physics, Philosophy of religion, and Diogenes. In her spare time she enjoys painting, eating frozen yogurt, playing devil's advocate and travelling. She hopes to one day become a Polymath.



ABOUT NICK RIGGLE, PHD



Nick Riggle is Associate Professor of Philosophy in the Department of Philosophy at the University of San Diego. His work centers around questions about the nature of aesthetic value, its place in our lives, and its importance for our practices and institutions. He draws on and contributes to work in value theory, philosophy of language, history of philosophy, moral psychology, social philosophy, and other areas. He also writes for non-academic audiences and has published work in various popular outlets, including two books: *On Being Awesome: A Unified Theory of How Not to Suck* (Penguin Books), and *This Beauty: A Philosophy of Being Alive* (Basic Books). His work has appeared in *The Philosophical Review*, *Mind*, *Philosophy and Phenomenological Research*, *The New York Times*, *McSweeney's*, *LitHub*, *Aeon*, and others outlets.



STANCE: We really appreciate you meeting with us today. We've been reading through pretty much all of your work, and we've been really enjoying it.

RIGGLE: Yeah, happy to hear that.

S: We wanted to start by giving you the opportunity to talk about your educational journey. You've had a bit of a nontraditional path. If you'd like to, just give us a quick rundown of the path of your education and your professional career, and touch a little bit on how participating in the X Games shaped your philosophical outlooks.

R: I never really cared about school until I was twenty or twenty-one. That was largely because I fell in love with rollerblading—street inline skating—around the age of eleven. Instead of being at school, I was at the skate park. And it didn't help that when I got to high school at age fourteen, the skate park was literally across the street. This was in Santa Rosa in the '90s. So, I was just really distracted. I didn't care about doing my English and math homework or passing a geometry exam. School just couldn't matter to me, because this other thing mattered so much. I was deeply in love with skating and devoted all of my extra time and energy to it.

My household wasn't super stable and being at home wasn't always the most enjoyable thing. And I had that other incentive to be at the skate park and hang out with people who I could look up to and emulate, people who were really cool, talented, funny, and supportive. By the time I was fifteen or sixteen, I was getting a lot of real opportunities to travel, make money, and compete internationally as a professional.

But my parents were supportive of my skating. They didn't ever have much money, and here I am getting a salary and plane tickets to Japan, Paris . . . all kinds of places throughout the world. They were supportive. They were like, "Hey, we can't provide this for you, and you're doing good things, so you should pursue it."

My high school wouldn't recognize the legitimacy of my sport. They would recognize ice skating and gymnastics. And they would offer independent study opportunities for a child actor or someone who is talented in some recognized sport. So, I ended up dropping out my junior year.

I was seventeen when I fell in love with reading. It's funny, when people were telling me what to read, I just had no interest. But when I could read whatever I wanted, I realized I had this voracious appetite for studying. I was interested in everything. I was reading



books on world religion, I was reading all kinds of fiction that I had never read, I was exploring philosophy and studying music theory—I was just all over the place. An upshot of dropping out was realizing my passion for learning.

BUT AS A
TEENAGER, AND
ESPECIALLY AS A
YOUNG ADULT,
I DID WANNA
DETERMINE MY
OWN VIEWS.
I WANTED TO
FIGURE IF I BELIEVE
IN THIS STUFF
OR IS THERE
SOMETHING
ELSE THAT I FIND
PLAUSIBLE?

I eventually found my way to philosophy. I kept coming back to it. I grew up in a Christian house, and my mother is very religious. I never identified with the religion, it always felt very strange to me. I didn't understand the tenets or the stories, and I just I didn't connect with it. But as a teenager, and especially as a young adult, I did wanna determine my own views. I wanted to figure out if I believe in this stuff or is there something else that I find plausible? I fell in love with Buddhism, I studied it very intensively for about two years. Both in theory, but also mostly in practice, I was obsessed with meditation. It was actually through meditation that I realized that my true love was philosophy.

I thought maybe Buddhism would answer some questions I had about existence, meaning, and spirituality. I ended up being really skeptical of some of their metaphysical claims, philosophy of mind, and theory of consciousness. A lot of it was very interesting, but it wasn't something I was just gonna swallow wholesale. And my philosophical, skeptical mind eventually recoiled. I found, "Oh no. Philosophy is where you actually develop this more critical approach, this more argumentative and logical approach."

By the time I was twenty, I was thinking about developing my career in extreme professional sports, which would be starting a wheel company or becoming a tour manager—skaters can't really be a top pro forever. People expand, they build out their career with other things. I was looking at that. But I also realized that I could go to community college. I dropped out of high school, but I could still go to community college and see what's up with studying philosophy. I didn't really have the stomach to be the owner of a company or do something within the industry. It didn't feel right to me.

I kept skating, but I also started taking philosophy classes at the community college and never looked back. I did really well in



community college. I was able to transfer to UC Berkeley, and I did really well there. I was just laser-focused on philosophy—the rest is history.

S: That's awesome. I think that would connect with a lot of people who don't have a traditional path to any field of study. This feels like a great spot to jump into the area of philosophy that you focus on. Most of your writing, especially recently, has been about beauty, and aesthetics more generally. Was there any pivotal event or text that inspired you to take beauty seriously as a philosophical topic?

I'VE ALWAYS
DEEPLY CARED
ABOUT BEAUTY,
SOMETHING
WHOSE POWER
OVER ME SEEMED
SO MYSTERIOUS,
AND SOMETHING
WHOSE VALUE IS
TOTALLY UNCLEAR.
AND BEAUTY
IS RADICALLY
UNDER-THEORIZED
RELATIVE TO OTHER
MAJOR CATEGORIES
OF PHILOSOPHICAL
INQUIRY.

R: I had a funny experience when I was at Berkeley. I was a transfer student. I was there for five semesters and then hung around for a semester to TA for a couple logic classes. When I was there, I really focused on philosophical logic, philosophy of language—these more austere technical branches of philosophy, which I still enjoy. But, every summer, when I wasn't doing assigned work, I kept obsessing about beauty. If you were to ask me, "What do you think about over the summer?" I would say "The nature of beauty." I've always deeply cared about beauty, something whose power over me seemed so mysterious, and something whose value is totally unclear. And beauty is radically under-theorized relative to other major categories of philosophical inquiry.

But I was unable to study beauty at Berkeley because they had no classes in aesthetics at the time. I was faced with that gap in my education. When I got to graduate school, that gap persisted. There was no one at NYU—where I was a graduate student—who even offered undergraduate classes, let alone graduate seminars, in aesthetics.

In graduate school I was pretty dead set in my first couple of years on exploring aesthetics. In my second year, I noticed that Noël Carroll was offering a series of seminars up at CUNY. The fall seminar was Classics in Aesthetics and the spring one was Contemporary Aesthetics. It was pretty ideal, because CUNY was



just right up the street from NYU, and I was able to sit in on those seminars. I took them for credit and learned a ton.

I'D TRAVEL AROUND
THE WORLD AND
SEE STREET ART
EVERYWHERE. WHEN
I WENT TO NEW
YORK IN 2007, THERE
WAS A STREET ART
BOOM, THIS POST-
9/11 FLOURISHING
OF STREET ART
EVERYWHERE. I WAS
OBSESSED AS A FAN.

In fact, for the spring seminar I wrote the paper on street art that eventually was published in 2010, which is my first publication! I asked Noël Carroll, “Hey, what should I write on?” I had two ideas. One of them was on political art or something; I don’t even remember what it was. But the second one was on street art, which I’d been a fan of since my skating days. It was all over the skate parks, and I’d travel around the world and see street art everywhere. When I went to New York in 2007, there was a street art boom, this post-9/11 flourishing of street art everywhere. I was obsessed as a fan.

When I did research about street art, I couldn’t find anything. There was some stuff, but nothing in philosophy. Arthur Danto published a couple of essays on early graffiti in the ’80s, but they were art criticism, not so much philosophy. I pitched this idea to Noël Carroll and he lit up. He was like, “You should write that; that would be great.” I did, and I had a blast writing it. He commented, “You should send this to the *Journal of Aesthetics and Art Criticism*.” I was like, “What? What even is that?” I worked on it little bit, sent it to the journal, and a year later it was published. That whole experience—the joy that I felt in writing that paper and the coolness and welcoming nature of Noël Carroll—was something that really got me thinking, “Oh, maybe I should work on aesthetics at NYU. No one is doing it here.” I went for it. I was working with a few people who were really open to it: David Velleman, Béatrice Longuenesse, and Paul Boghossian. Ultimately, near the end of my time there, NYU hired Rob Hopkins, who is a genuine specialist in aesthetics, and very talented. I got his advice towards the end there, and it all worked out. It was a bit of a leap of faith, really, to specialize in.

S: That’s really cool. This seems like a great spot to segue into a question we intended to ask a little bit later. You mentioned that there was almost a total lack of opportunity for you to get invested in aesthetics. Did this have any effect on how accessible your work is? When people think of philosophers, they think of thick language like Kant and other old-timey philosophers. But your work is very



accessible. You use a lot of everyday examples. Did your experience accessing aesthetics in graduate school have an effect on how you write your books?

R: That's a great question. One of the main things is that I love writing. I think I would be a writer no matter what career I went into. And I don't just write philosophy—I recently finished a novel that I'm trying to sell. I've written poetry for almost my whole life. I don't share it, really. I was in a band for several years and wrote a lot of music. I just love writing. I write lots of different kinds of things and just enjoy the process.

Gosh, almost twenty years ago when I was in graduate school, there were a lot of expectations to fit your academic writing into a certain formula. I think there's been a lot of progress in professional philosophy where the formula of the academic essay is a bit looser. But back then I found, and still do find, the formula for academic writing utterly oppressive. As a person who really cared about creative writing, when I was a graduate student attempting to respond to these demands, I needed an outlet. I was feeling my creative side wasn't flourishing. So, in graduate school, I was writing a bunch of stuff. I started working on fiction more seriously and poetry was an outlet.

At one point, I really got lucky. I wasn't really thinking about writing trade books in philosophy. It hadn't really crossed my mind. But, as I was finishing my dissertation, I had this idea about the nature of awesomeness. In fact, it started as an idea about the nature of suckiness. My wife and I were just joking one day and thinking—I hate to say it—but some people just suck. “What is that? What is that quality?” We were just puzzling over it. It just seems true, and it's not clear what exactly that picks out. As a philosopher, of course, I'm like, “Hmm, this is really interesting.” I quickly noticed the opposite of suckiness is awesomeness, so, what's that?

As a parenthetical, since you're students, I'll give you this rule that I've followed my whole life: every idea gets five minutes. You have to be attentive to any idea you have. When you have an idea, make note of it, and when you have time, spend five minutes developing it. Write about it, think about objections to it, and do a little research on who's already written about it. What you'll notice is that over time, the good ideas snowball. Once you've spent five minutes on a good idea, then later on in the day, later on in the week—at some point—you'll have another idea about that idea, and that gets five minutes. And the bad ideas: you write them down, and they don't snowball; they don't come back to you.



So, this idea about awesomeness snowballed, and I kept thinking about it. Then I got lucky again, because right as I had a mini-essay and a little outline for this series of ideas, Aaron James visited NYU for a semester or a year. Aaron James had recently come out with a book, *Assholes: A Theory*. It's in the same spirit—it's a trade book. It sold really well, and it's about what it is to be an asshole. I asked him out for lunch and pitched him on the idea about awesomeness, and he loved it. I remember very vividly, he said, "This idea has legs." Meaning, it's gonna go somewhere. That just encouraged me.

I FELT IN MY HEART
I HAD TO TRY TO
COMMUNICATE
WITH A BROADER
AUDIENCE,
PEOPLE WHO I
FELT MORE OF AN
AFFINITY WITH.
IT WAS ALMOST
LIKE A WAY OF
STAYING TRUE TO
MYSELF AND MY
BACKGROUND.
WRITING TRADE
BOOKS AND PUBLIC
PHILOSOPHY
SPEAKS TO THAT.

I was on the job market that year, which was really stressful. Working on the book gave me a lot of joy that I needed. I needed a counterweight to all the stress I was under to try to get a job in academia. That encouraged and influenced me to write on this idea even more. One thing led to another, and it became a book. I think part of it is that I have an obsession with creative writing, and then the other part is that I'm a first-gen student who was a pro skater. I just think my personality is very not academic. I don't fit in that well. I felt like I could communicate with academics, but also with non-academics. I felt in my heart I had to try to communicate with a broader audience, people who I felt more of an affinity with. It was almost like a way of staying true to myself and my background. Writing trade books and public philosophy speaks to that.

*S: That's awesome. It reminds me a little bit of what you talk about in being a creative community builder. It sounds like your work is a form of that. To jump into the specifics of your book *On Being Awesome*, why don't you just give a brief overview of what it means for someone or something being awesome versus sucking?*

R: Being awesome is being good at creating social openings. Sucking is a matter of declining to take up social openings for no good reason. A social opening is an opportunity to break out of our social roles and our habits and express our individualities to each other in conditions where we don't always get to. When we're out in public, we're often just enacting a script that's more



or less determined by a social role—that of a customer and an employee, or a person walking down the street. When enacting these social roles, we don't have terribly many opportunities to just be ourselves. A social opening is an opportunity to break out of that role, to put those habits and routines aside, to express individuality, your sense of humor, your taste in food and drink, your sense of playfulness, your sense of what's good in music, and so on. You can think of a range of aspects of individuality that you might express given the opportunity. So, awesome people are good at noticing when there are these opportunities—taking them up or creating them. Sucky people are ones who decline the opportunity, essentially for no good reason. They notice it, but they just can't be bothered.

S: Right, perfect. That's generally what we gathered. Staying in the same vein, another concept that you talk about is volitional openness. Our understanding of this idea is that it means someone's open to opportunities of awesomeness; that they're receptive, they're open to new ideas or experiences, and that they don't just turn them down for no good reason. Could you expand a little bit on how volitional openness relates to awesomeness and suckiness?

R: Good. In *On Being Awesome*, I don't use the concept of volitional openness, but I do use the concept of play. The philosopher that I'm inspired by with respect to volitional openness, is Friedrich Schiller. His notion is playfulness. Play, for him, is what really matters. It's not just being open to ideas; it's actually a mode of engagement with value. Volitional openness is a way of being receptive to, responsive to, and engaged with the values that are immediately present to you in your environment. Its contrast or complement is autonomy. As autonomous beings, we have our values. We have the things that we've committed to, the responsibilities we have taken up, the promises we've made. And those are self-shaping. Those are important. They make us who we are as autonomous beings. But—and this is really Schiller's point—autonomy limits freedom. Although autonomy allows us to be free,

AS
AUTONOMOUS
BEINGS, WE HAVE
OUR VALUES.
WE HAVE
THE THINGS
THAT WE'VE
COMMITTED
TO, THE
RESPONSIBILITIES
WE HAVE
TAKEN UP, THE
PROMISES WE'VE
MADE. AND
THOSE ARE SELF-
SHAPING. THOSE
ARE IMPORTANT.



in a Kantian sense, from our inclinations, it also constrains us, because it can limit our openness to value.

THE PLAYFUL PERSON
IS NOT GONNA
FORGET WHO THEY
ARE—THAT WOULD
BE MORALLY BAD—
BUT THEY’RE ABLE
TO TEMPER THEIR
AUTONOMY, QUIET
IT DOWN, AND SET IT
ASIDE. THE PLAYFUL
PERSON IS ABLE TO
JUST BE IMMEDIATELY
RESPONSIVE TO
PRESENT VALUE: THE
CONVERSATION,
THE SCENTS, THE
SIGHTS, THE SMELLS,
THE SOUNDS, THE
PEOPLE.

Think of a lawyer who is very into being a lawyer. A lot of their dispositions and habits and thoughts are just lawyer thoughts and dispositions and habits. They come to your dinner party and they’re just a lawyer being a lawyer. They’re always responding in a lawyerly way, bringing up lawyer things, and being very “lawyer-y.” Schiller would say their form drive is too dominant. He would say, “They’re not playful enough. They don’t understand the value of playfulness.” The playful person is not gonna forget who they are—that would be morally bad—but they’re able to temper their autonomy, quiet it down, and set it aside. The playful person is able to just be immediately responsive to present value: the conversation, the scents, the sights, the smells, the sounds, the people. When you’re like that, Schiller thinks, you’re at your most free; that’s the highest freedom that you can possibly attain.

Schiller also thought that beauty and art were the things designed to actually help us enter that state of volitional openness, that state of play. Beautiful paintings with their vivid colors—boom! All of a sudden, we’re just immediately attuned to the values in front of us. And we’ve forgotten about our lives. You go see a movie, and then it ends, and the lights come on, and you’re like, “Oh wait, wait, oh—oh gosh. I’m here, in the world. I’m back. What?” You were gone. You lost yourself in the immediacy of the present experience. So, Schiller says beauty is the only path to freedom because it’s the only thing that you can regularly bring you to this ultimately free being. Beauty is actually able to temper autonomy and open you to immediate value in this volitionally open way.

One addendum here. Playfulness does feature in *On Being Awesome*, but given the description of volitional openness, you can see that openness to social openings requires playfulness. Again, if



someone gives you a social opening in the middle of your day, you could just continue to be the autonomous being you are and just ignore it. Or you can notice that opportunity, be volitionally open and responsive to it, and be playful.

S: Perfect. As a point of clarification: How do we differentiate between when someone is given an opening and rejects it (therefore being sucky) versus someone who rejects an opening as a matter of taste or preference?

R: I think it's a little delicate. I think there are a lot of ways of rejecting a social opening as a matter of taste that are also sucky. I think good aesthetic interactions and social aesthetic nature requires two things: (i) the vulnerability of expressing yourself as an individual, putting yourself out there as someone who has this taste in music, who has this taste in painting, who likes these things and not those and (ii) an openness to the expressions of other people you're interacting with who are also being vulnerable in that way, which is where volitional openness and playfulness come in.

IT'S A DANCE
BETWEEN
VOLITIONAL
OPENNESS AND
WHAT I CALL
DISCRETIONARY
VALUING.

It's a dance between volitional openness and what I call discretionary valuing. You're trying to do both in a harmonious and supportive way. That's aesthetic community for me. Awesomeness really is just a different word for aesthetic community. If you're rejecting a social opening because it's not your taste or it doesn't fit your individuality, well, that's a bit sucky, because the whole idea is to be open to expressions of individuality. But, that said, tastes are wild—some are incompatible. It's reasonable sometimes to decline a social opening, because there is no productive vaesthetic community there. You don't find it interesting to engage with someone who really loves that music. I think that's totally possible. To give you specific examples is hard. But surely that's possible.

S: Yeah, that makes sense. I think that gives me a good idea of where the line is. Just a minor question: Would you say that noticing awesomeness is a skill that's built up from experience?

R: Absolutely. My developed view about the nature of aesthetic value and aesthetic community is that it's a skill. Part of that is this



idea that we have these basic and general capacities. Volitional openness and discretionary valuing are general human capacities, ones that we're born with. I'm open to that. The capacity is innate or natural. But aesthetic valuing calls on us to do a couple of things. One is to take these two capacities that are independent and exercise them together. We're asked to be playful individuals, volitionally open to discretionary valuers and to do that socially. To yoke them together in yourself and then express yourself in community with others takes skill. The artistic and aesthetic media we deploy to socialize around these capacities takes skill to engage with, to create, to share.

S: Great, thank you. To stick to awesomeness and suckiness, you talked about how the initial idea for this book actually started on suckiness and then developed into awesomeness. We were wondering could there be awesome people if sucky people didn't exist?

I THINK WE COULD
RECOGNIZE
THE RICHNESS
OF DIFFERENCE
IN AESTHETIC
VALUING AND
APPRECIATE
THE VALUE OF
AESTHETIC
COMMUNITY
WITHOUT THERE
BEING VERY MUCH
SUCKINESS, OR
MAYBE NONE AT
ALL.

R: I think it might be harder to notice, but I think it's still possible. To put a little more meat on those bones, part of it is discretionary valuing, which is essentially choosing what to care about in your aesthetic life—choosing what music to love, what fashion to embody, what cuisines to geek out over, and stuff like that. Different people are gonna exercise that capacity in different ways, and I think there's a lot of opportunities to engage with one another productively and aesthetically without being sucky, without being dismissive, rude, or gratuitous in your dislike of certain things. So, I think we could recognize the richness of difference in aesthetic valuing and appreciate the value of aesthetic community without there being very much suckiness, or maybe none at all.

*S: Right. This feels like a great place to talk about a connection in your other book, *Aesthetic Life and Why It Matters*, that you wrote with Bence Nanay and Dominic Lopes. In the final breakout chapter, in the section on individuality, you say we waste our time on bad art, ugly clothes, and sucky bands. Is there a connection between aesthetic objects individually and your theory of awesomeness? Are there objects themselves that are awesome for the psyche?*

R: Right. One way to think about being awesome in my theory of aesthetic value, is its magnified look at aesthetic socializing. What



does that framework look like when it's just put into the wild? How do we engage with these values as social beings? Basically, my theory of aesthetic value is that the things that have aesthetic value are the ones that are worthy of aesthetic valuing, and those are the things that are capable of grounding aesthetic community which, for me, is the mutually supportive exercise and cultivation of the capacities for discretionary value and volitional openness. So, aesthetic community is actually a process. It's a thing we do together. Aesthetic objects, things that have aesthetic worth are things that are conducive to awesomeness. Like, a good joke could be aesthetically good for the same reasons, because it can create a social opening. And the product of a successful social opening is aesthetic community. In the book I call that awesomeness. This value that we seek in being awesome—it's the same. It's essentially the same thing as an aesthetic community.

It's a little funny. In a way, *On Being Awesome* should have been written later in my career since it is basically a social application of more foundational works I published later. Writing it helped me articulate a bunch of ideas I had about the nature of aesthetic value. But it's only recently that I have published the main paper that gives the whole theory, "Aesthetic Value and the Practice of Aesthetic Valuing."

S: As a follow-up, you're talking about what can make an object aesthetically worth valuing. In that same breakout chapter, you mentioned that you don't think that individuals can make legitimate demands on anyone else to aesthetically value in a particular way or another. Is there a way for us to make clear and homogenous determinations of what is worth aesthetically valuing? Or is that necessarily a subjective lens that always comes into play?

R: My view there has to do with the nature of discretionary valuing. Remember, aesthetic valuing for me comes first, because aesthetic community is the highest good in aesthetic life. Aesthetically valuable objects are a means of grounding aesthetic community. It's part of what's essential to aesthetic community is this capacity for discretionary valuing, an individual capacity to choose for yourself what to value, how to value it, how to interpret it, how to be inspired by it, and how to incorporate it into your life.

I don't think anyone can tell me how to exercise that capacity. What you can do is invite me to value something. I think an invitation for me to exercise my aesthetic valuing capacities on the thing is fine. I develop this idea in a paper called "Convergence, Community, and Force in Aesthetic Discourse." In my view, when you say to me that



something's beautiful, socially, what you're doing is attributing a certain aesthetic property to it and then inviting me to value it. You're not necessarily saying that I must agree with you that the object has that property. You're just saying, "Hey, Nick, tell me what you think?" Aesthetic claims have the illocutionary force of invitation.

I don't think people can make demands about what I ought to value aesthetically. But I think we can make sense of the idea that, for some people, certain things just are not worth their time. But that has to do with facts about the connection between our ability to exercise these capacities and things like our circumstances, our personalities, and our other goals in life. It might be that I'm exercising my discretionary valuing over a certain band, but it's partly because I'm in a certain social group. I'm gonna really try to like it, but it's just not my personality. There could be a clash between personality and an aesthetic item, there could be clashes between your circumstances and the time you might spend doing a certain thing. It's stuff like that, which would generate potential wastes of time.

S: Personality segues nicely into our next main topic: artistic personality. We were wondering: Is artistic personality something that exists across all art forms, even ones that might be conducted alone? For example, you mentioned that you don't really share your poetry. Another example is the painter Francisco Goya, who painted a lot in his home with no intention of the paintings ever being shared. After he passed, they were discovered. Where do those fit into your concept of artistic personality?

I THINK THE
CULTIVATION OF
AN AESTHETIC—
SPECIFICALLY
AN ARTISTIC
PERSONALITY—IS
SOMETHING THAT
WE CAN DO ALONE.
BUT NOTICE THAT
WHENEVER YOU DO
IT, YOU'RE ENGAGING
IN A TRADITION.

R: In a couple of papers, including a co-written paper with Rob Hopkins, artistic style is the expression of ideals—ideals that you have for your own art, as an artist. I think you can develop those ideals on your own. I think it's typical that people develop them on their own. I think the cultivation of an aesthetic—specifically an artistic personality—is something that we can do alone.

But notice that whenever you do it, you're engaging in a tradition. You're responding to a practice that has existed for a long time. So, I'm not so sure that people can cultivate an artistic



personality on a lone island. It's in response to the painters that have come before you that you develop your own painterly ideals. It's in response to the filmmakers who came before you that you can develop your directorial ideals, and so on. You might, like Goya, create a work that no one ever saw in his time. Maybe he was even thinking, "No one's ever gonna see this." Does it still have aesthetic value? My answer is yes, because he's engaged in aesthetic practice, which is a social practice. Once the social practice exists, well, people can engage in that practice in all kinds of deviant or normal ways.

Still, the products they make are products of the practice. If they're successful or if they're good products of the practice, they'll have the value that that practice is designed for, creating aesthetic value in this case. One analogy that I like to use is baseball, which is an enormous communal social practice. It involves fans, coaches, umpires, and teammates who have all kinds of different roles, and they have to work together. You can still practice baseball by yourself. You could throw the ball up and hit it. You could throw it against the wall. There's all kinds of things that you could do by yourself. But when you do, you're still engaging in the practice of baseball. You're just doing it alone. Of course, there are limits to that, but I think once you have a social practice, there's all kinds of ways you can develop skills for it, contribute to it, or act somewhat deviantly towards it.

S: That circles back into an example you talk about: the origins of the high five, that it actually coming from baseball. You talk about it as a one of the practices of being a creative community builder. Is there a specific favorite way for you to be a creative community builder? Do you like the high five best or are there potential practices that you tend to lean towards?

R: I do love the high five. I would say the big picture is food. I'm really chef-y. I love cooking. I've been doing it for years. I'm always experimenting with new things. I love cooking for my friends; I love having them over for dinner, creating that social opening, the dinner party, the barbecue—whatever it is. Food is my thing.

*S: I also love cooking. That leads me to a question about something you said in *Aesthetic Life and Why It Matters*. You talked about how food is the stuff that's worth eating, that's worth valuing. You mentioned some examples, including things like chichitanas. But I noticed you didn't have an example from your life. Is there a specific dish that is your go-to ground for community and for aesthetic valuing?*



R: Yes, wine. I make wine. I have a giant 700-liter terracotta amphora from Italy that I use to make wine in with my friend Chelsea Coleman. Wine making is a wonderful practice. Every year there's a harvest from mid-August to mid-October. We help each other out, and we get our haul. The amphora takes one ton of grapes, which would basically fill up the back of an F-150, pile it up over the top and you've got a ton.

We harvest our grapes and crush them. They go through primary fermentation and we spend weeks tending to them, making sure everything's developing right and smells all good. Then they go through *élevage* for however long—that's the second phase where the wine is aging. At some point, you press the grapes, probably age it a little more, then you bottle, and you drink it with your friends. The whole process is just a joy. It's labor, it's sensory, and it's beautiful. So, making and sharing wine is a big deal for me and creates a lot of social openings. I would say it is super awesome.

S: We have a few more questions. In "On the Interest in Beauty and Disinterest" you talk about the concept of disinterest and how it has misguided people. If the importance of beauty becomes less profound, what do you think we would lose? Would we lose some aesthetic values towards beauty?

I THINK IT WOULD BE A SHAME IF, AS A CULTURE, WE LOST TOUCH WITH THOSE SELF-INVOLVING, LOVE-INVOLVING, PASSION-INVOLVING ASPECTS OF BEAUTY. I THINK WE'D LOSE TOUCH WITH BEAUTY AS A SOURCE OF INSPIRATION AND COMMUNITY.

R: In one way, the concept of disinterest is very academic. It has obviously influenced a lot of philosophers to think about the nature of aesthetic affect, though I think a lot of people in the wild don't really think of beauty that way. They think of it as passionate and erotic, involving love, self, and profundity in ways that are not well-theorized by the concept of disinterest. I think it would be a shame if, as a culture, we lost touch with those self-involving, love-involving, passion-involving aspects of beauty. I think we'd lose touch with beauty as a source of inspiration and community. I think there's a lot of community in those features of beauty.

S: You mentioned briefly the poser: someone whose ideal is a certain set of ideals. We were curious to know how that relates to technology and social media, and



how algorithms are giving the ideals to you? Do you think social media and algorithms are adding to disinterest, making people think about beauty less?

R: It's hard to say. I do think there are a lot of algorithms that are designed for engagement. If they're working, then they're engaging people. That's their whole purpose, because they're money makers. The more engaged people are, the more ads are sold, and the more money is made. I think the problem comes from the fact that [algorithms] influence us almost without our knowing it. And in that way, they diminish our capacity to exercise our discretionary valuing and undermine aesthetic practices.

IF YOU MAKE
THE EFFORT
TO CULTIVATE
YOUR OWN
INDIVIDUALITY AND
EXERCISE THAT
DISCRETIONARY
VALUE IN A WAY
THAT'S TRULY
YOURS, I THINK
SOME OF THESE
TOOLS CAN BE
QUITE LIBERATING.

But I also think I don't wanna be too pessimistic about this. I think some of these things are actually quite empowering. I've been introduced to a lot of new music, and I'm able to find obscure films and find communities of people who are talking about these films they love. So, I'm a bit on the fence. I think the dangers are real. I think they're undermining aesthetic practice, especially for young people. But I think if you make the effort to cultivate your own individuality and exercise that discretionary value in a way that's truly yours, I think some of these tools can be quite liberating.

I'm working on a new trade book right now exactly about this. One of the arguments is that we're no longer in the age of criticism. Criticism was a very twentieth-century thing. If you wanted to engage with criticism about literature, film, painting, or whatever, you'd read long-form criticism in periodicals and newspapers. People just don't do that anymore. Instead, they look at lists: the top ten books of 2020 or the top twenty albums of the twenty-first century. You might have an "Ugh, they're just trying to sell more stuff" cynical attitude toward lists.

But I actually think that it's not just publishers and media companies giving us these lists. If you go to TikTok and Instagram, they're scorers. "Oh, it's 9.6 out of 10, yeah." I think we're in the age of lists and scores, and this age is under-theorized. I think that we can engage with lists and scores in a really cool, productive way for the practice of aesthetic valuing. Although there's some reason



to be a little cynical, I think there's also room for an apology or defense of lists.

S: That really interesting. A lot of times this list-making is an aesthetic practice that builds community. Someone makes a list, then a bunch of people agree or disagree about whether its rankings are appropriate.

R: That's exactly right. I think it's an aesthetic practice that has replaced criticism, the predecessor practice. A list is really an invitation, and you can see it play out online. Someone gives a list on TikTok. Then comments open up, Boom! "Oh, but what about this movie?" "I think four should be five because of this . . ." And there are people who don't get the invitational aspect of it, who think it's an epistemically strict list. They suck, right? They're not getting the invitation. They're dismissing it for no good reason. So, I think "the list" is very under-theorized. We're a quarter way through the century. We need to reckon with lists.

S: So, some people disagree with the ranking in the list. But what about a person who just wants to be a person who reads the top one hundred books? Is this person a poser? How does the idea of being a poser play into the idea of lists?

R: I don't think disagreeing with the list is bad. It's the register of disagreement that matters. It's dismissiveness that's the problem, as in "How could you put that at number one? You're an idiot." That's not good.

But I do think it's amazing if someone comes out with a list that you totally agree with. I play this game with Matt Strohl, a friend of mine—a really good philosopher and an incredibly talented film critic. We each have a ranking of all eleven *Fast and Furious* movies. If one day our lists converged, that would be cool. I'd bet they don't converge. We have disagreements, but all by way of aesthetic community.

We can diagnose the problem with being a poser within my system as a failure to understand the nature of aesthetic community. If you want to like something just because other people like it, you might get community of a sort, but it's not necessarily aesthetic community. It's actually a response to aesthetic value that fundamentally misunderstands the nature of aesthetic value.

S: To go a little back a bit into the algorithm stuff, we're really curious about how to make sense of the brain rot that comes from the stuff that algorithms feed us. Our scrolling isn't really an expression of our ideals or artistic sense. It's just fed to us and there's not a lot of passion behind it . . . To stitch that together with aesthetic merit, can we view something like that as having aesthetic merit,



even so much to say that it's beautiful; not in the sense that it's beautiful for us engaging with it, but from a third-person point of view. If a space alien saw that the humans were battling with brain rot, could it be viewed as something beautiful from that zoomed-out lens?

R: That's a really cool question. That takes a lot of theoretical imagination. Is the whole thing beautiful? Oh man, I have a million thoughts. I'm tempted to take this even more big picture. Human life itself is something we can zoom out on and see that as beautiful because it's full of chaos and struggle. I deal with this a bit in my book, *This Beauty*. It's structured on the idea that life is really difficult and full of pain. Life's difficulty is inescapable: the pain, the challenge, and the difficulty. And, of course, we die at the end, and no one asked to live, feel pain, or struggle. This is not to diminish the fact that we also feel joy. But the fact is, we didn't agree to this thing and we have to deal with it.

One response that people have had to this problem is a theodicy. You could justify existence with God. You could focus on the fact that the universe itself is beautiful, the universe is fundamentally good. You are made in the image of a loving being, God. It's a way of saying that everything's beautiful.

I CAN TELL
THEM THAT
LOVING
BEAUTY IS
A WAY OF
KEEPING IN
TOUCH WITH
THE VALUES
THAT MAKE
LIFE WORTH
LIVING.

I can't accept that. I'm a father and I can't tell my kids that. I can't. I don't think I would be honest. What can I say is that they didn't have anything to say about it. It was my choice. And they're gonna suffer and struggle, and I'm gonna die. And then they're gonna die. In the book, I developed the thought that I can tell them that loving beauty is a way of keeping in touch with the values that make life worth living. It's, "Hey, just make sure you're doing this a lot. Make sure you're in love with the beautiful." Because then you're gonna be developing your playfulness, your individuality, and you're gonna be in community with other people who are doing the same, and they're gonna support you. Next, you're gonna have these answers about why you should continue to live your life, even though you didn't decide to have a life and it's full of struggle and pain and suffering.

Let me zero in a bit more on brain rot. I'm a doom scroller. If what you're talking about is just scrolling through all the content,



that's actually a really unique, technological way to engage in the practice of aesthetic valuing. I do this as well. I mean, I have a long day, I need a reason to shut my brain down, and this is it. But in the process, I'm like, "Oh, this is funny" or "This is interesting" and I'll send it to my friend. This is a way of engaging in the practice of aesthetic valuing. It's not always aesthetic, it might just be information sharing, but a lot of it is.

S: Do you have a favorite brain rot trend at the moment?

R: I don't know. Maybe I'm too old, but I just find it all delightful and curious. I don't feel like I'm a super insider in a way. I would never be the one who created this stuff. But I don't know, I'm kinda open to it all.

S: What are some other fun things for us to talk about? You have a lot of interesting stuff but we're running low on time. Let's wrap up on this: You say there is an existential imperative of carpe diem. As you explain it, it's not about "seizing the day." Rather, it's about harvesting and plucking the day, of paying attention to what is around you, and interacting with the beauty that's presented to you. So, we were wondering, what have you done today to pluck your day?

R: Oh, man, well, it's only 12:21 p.m. Besides talking to you, the main thing I did today was go to the gym. I had a thirteen-hour day yesterday. I worked myself to the bone. So, I thought I was gonna get this really great night of sleep. Well, I woke up at five a.m. I was like, "What the hell?" On a Wednesday I normally would go to the gym, and I really didn't want to, because I was tired. But in the spirit of seize the day, I made it to the gym. I didn't just do a minimal workout. I went for it. So, that's about all I achieved today. Well, I also got the kids ready for school, made them lunches, etc. But I feel like hanging out with you guys is seizing the day! This is great.

S: We hope it's seizing the day! I have one more question. I'm pretty sure it was in On Being Awesome that you said swimming sucks? Why does swimming suck?

R: First of all, I'm probably not being fair to swimming. But to qualify that a little bit, what I wanted to say is that swimming is full of value. It's a great sport, it's just that it doesn't have the awesomeness-suckiness structure of value. Its value is end-oriented, it's about competition and achievement.

S: So, swimming doesn't compare to baseball or another team sport because it doesn't have the community aspect that you talk about?



R: I don't want to say it doesn't. There's a relay sport in swimming, which is team oriented. Rather, it's that there are these impressively strong and talented people going super-fast in water, and what matters is just being the fastest; it's very end oriented. I do think swimming is awe-inspiring in the classical sense. When you see what Katie Ledecky does, you can't help but think, "How?" It's so awe-inspiring. But it's structured around individual excellence at achieving a specific goal. So, there's no fundamental emphasis on things like cultivating your individuality or achieving a style. Again, maybe someone who knows more about swimming will put me in my place. I'd like to be put in my right place on this.

More generally, I do think there are different categories of value. I mention swimming just as a contrast case to things like rollerblading, skating, and skateboarding. Although there are skate competitions, every skater would admit that they're a little artificial. They're not the heart of the sport. The heart of the sport is cultivating your individual style in a community of people cultivating their individual styles, and you do it in a supportive and rad way. People wanna develop their own lexicon of tricks, they wanna do it with other people. Different people can do it in really different ways. It's not about getting the fastest time, jumping the highest, or whatever. It's way more communal and organic.

As a contrast, you might think of a sport, whatever it is, that's about just about meeting some goal in the best way, where cultivating individual style is not as central to it, and even, in some sense, being in a supportive community is not essential to it.

S: That makes a lot of sense within the context of your theories. So, we can consider this a call for papers for someone to justify the aesthetic value of swimming.

R: You guys are gonna publish this, then I'm gonna get a bunch of hate mail, like from Katie Ledecky. That'd be cool. I'll take hate mail from Katie Ledecky.

S: I don't think we have any other questions for you, Dr. Riggle. I'd like to, again, extend a thank you. Thank you so much for meeting with us. We really enjoyed this conversation!

R: Yeah, this is my pleasure. Thanks for thinking of me for this!



SUPPORTERS

| MAJOR DONORS: \$100,000+

Ball State University

| CHAMPION CIRCLE: \$9,999-\$5,000

The American Philosophical Association
Dr. David W. Concepción

| SUSTAINERS CIRCLE: \$4,999-\$1,000

Dr. George Barker
Devon Chapmon
Rachel Crawley
Dr. Steven Humphrey
Mike & Jessica Reynolds
Cameron McNeely
Salesforce.org Foundation
Dr. Kate (nee Juli) Thorson
Univ. of Louisville, Philosophy Dept

| THOUGHT SMOKE CIRCLE: \$999-\$500

Chris Coy
Belkisa Hrustanovic
Trevor Krogman
Tim Pierz
Eric Roesch
Kristen Ruhl

| ORIGINALITY CIRCLE: \$499-\$250

Isaac Brooks
Daniel Cole
Tre Skye Graves
Tim & Susan Johnson
Dominic Martyne
Kalli McBride, JD
Caleb Mosier
Kiley Neal
Harrison Null
Dr. Gina Schouten
Marc Semanchik



SIGNIFICANCE CIRCLE: \$249-\$100|

Iesha Alspaugh
Lorrie Caskey
Craig, Kelley, and Faultless, LLC
Gatlin Freiburger
Sophie Gordon
Ashli Godfrey
Viveck Hadley
Debra Harvey
Emily (nee Hinman) Binedect
Richard & Janet Johnson
James Lacy
Sidney M. Lewellen
Kevin Mager

Owen T. Miller
Bailey Painter
Dr. Chin-Sook Pak
Whitley (nee Lewis) Pike
James Robinson
Josh Savage, JD
David Shepherd
Adam Vaughn
Dr. Sarah Vitale
Robert Wilbur
Kayla (Wynne) Williams
Ryan Zerfas

CLARITY CIRCLE: \$99-\$50|

Chris Baker
Samantha Carpenter
Dianne Crusan
Lauren Fosnight
Kyle & Aspen Johnson
Eric Lawler
Jeff Neff

Elizabeth Palmer
Ben Rogers
Dr. Jeanine Weekes Schroer
Bruce & Joyce Shepherd
Ben Slightom
Richard Storey
Caden J. Wiederspahn

I ♥ STANCE: \$49 OR LESS|

Gabriel Baker
Cecilia Becker
Chandler Bell
Jeremy Christman
Annamarie Clobuciar
Alisha Fields
Cecilla and Anthony Georgi
Kayla Gurganus
Tristan Hankins
Jason Holtzman
Rowan Hartje
Leah Heim

Matt Hinkleman
Michael J. Huerkamp
Ash Jefferies
Gabriel Keel-Brown
Waleed Ma'arouf
Sara Maier
Cory Matters
Michael McCauley
Ben McIntosh
Jack McKown
Zack Medford
Robert P. Mehling

Isabel Parham
Katherine Priddy
Cassandra Reed
B. Eliza Sandlin
Leo Silberhorn
Rosamae Swoape
Sid VanMeter
Daniel Wills
Esther Wolfe



