LOGOS

The Cornell Undergraduate Journal of Philosophy

LOGOS

Volume XXII - Spring 2026

Including:

Remedying the Malaise of Work:
A Dialogue Between Abraham Joshua Heschel and Josef Pieper
Elliott Jones - Boston College

Aristotle's Natural Slavery in Politics
Reconsidered Through De Anima and Nicomachean Ethics
Abigail So - Cornell University

Making Sense of Mental Ballistics
Uriel Vargas - University of Chicago

Volume XXII - Spring 2026

logos.philosophy.cornell.edu

“The stars went out,

few by few,

toward the west.

And still the family stood about

like dream walkers,

their eyes focused

panoramically,

seeing no detail,

but

the whole dawn,

the whole land,

the whole texture

of the country

at once”

Grapes of Wrath

John Steinbeck

Cover Artist: Darius Pierce

Logos : λόγος

Volume XXII • Spring 2026

CORNELL UNIVERSITY
SAGE SCHOOL OF PHILOSOPHY
Ithaca, New York

Logos : λόγος

Volume XXII : Spring 2026

An independent student publication

Logos: The Undergraduate Journal of Philosophy, an independent student organization located at Cornell University, produced and is responsible for the content of this publication. This publication was not reviewed or approved by, nor does it necessarily express or reflect the policies or opinions of, Cornell University or its designated representatives.

Mail:
Logos: The Undergraduate Journal of Philosophy
c/o The Sage School of Philosophy
218 Goldwin Smith Hall
Cornell University
Ithaca, NY 14853

Email:
journal.logos@gmail.com

Editor-In-Chief:
Corbin Breeden

Co-Presidents:
Julian Kanu
Wyatt Sell

Treasurer:
Marco Chong

Secretary:
Maeve McCarthy

Underclassman Representatives:
Annelise Durham Gross
Sammi Wu

Editorial Staff:
Sid Bajaj
Shula Brukh
Ella-Rose Cespedes
Louis Arnoult Costafreda
Alexander Coulouras
Jackie Dorland
Margaret Fear
Ethan Fisher-Dayn
Maria Yaniris Gil
Alejandro Ferrer Gomez
Jesse Green
Amelia Klaus
Kayla Lee
William Liaw
Maria Lim
Naveen Mathew
Jacob McCoy
Casey Platkin
Isabelle Schwartz
Robert Wallace Sims
Will Taylor
Hema Venkatraman
Richard Zhang
Emily Zhou

Faculty Advisor:
Harold Hodes

Contents

Editors' Introduction

The world is a strange and beautiful place. We constantly strive to understand this beauty and strangeness. Philosophy, in all of its forms, allows us to explore and appreciate reality. I hope the words you find in these pages help guide you through the labyrinth of life. The papers published here touch on topics from communication with multi-dimensional space aliens to how we should balance our busy lives, and of course, to the age-old philosophical discussion of free will. On behalf of the entire editorial board, it is my utmost pleasure to present the twenty-second edition of *Logos: The Cornell Undergraduate Journal of Philosophy*.

This year, *Logos* received 105 submissions from brilliant undergraduate philosophers around the world. Sixteen advanced into our final round. After careful deliberation, we found the six published here to be extraordinarily interesting, rigorous, and relevant. They each show a masterful comprehension of the relevant philosophical context while simultaneously pushing the bounds in their originality and creativity. We are extremely grateful to our authors for their contributions, and dedication to the craft of philosophy.

I would like to extend thanks to the editorial board for their constant hard work. It is no easy feat to juggle a full courseload, social commitments, and editing a philosophy journal! This journal would not be possible without their thoughtful insight, teamwork, and attention to detail. I would also like to thank the Sage School of Philosophy for their continued support of *Logos*. Pamela Hanna, Jacquelyn Johnson, Melissa Totman, Allison Barrett, and Ryan Belle have been infinitely kind, and made the process of creating, funding, and printing a journal incredibly easy. Thank you to Harold Hodes, our faculty advisor, for your continued dedication to *Logos*. Finally, the entire Editorial team would like to extend best wishes to Pamela Hanna. We hope you have a spectacular retirement, and greatly appreciate all you have done for *Logos* and Cornell Philosophy.

———

Corbin Breeden
Editor-in-Chief

Remedying the Malaise of Work: A Dialogue Between Abraham Joshua Heschel and Josef Pieper

Elliott Jones
Boston College

FOREWORD

Editors: Sid Bajaj, Shula Brukh, Margaret Fear, Alejandro Ferrer Gomez, Maeve McCarthy, Jacob McCoy, Casey Platkin

"The Puritan wanted to work in a calling; we are forced to do so." Max Weber.

The intertwined secular, individualist, and capitalist turns of Western modernity have wrought a new relationship between humans and work. This absence of religion and collectivism, reminiscent of Durkheim's anomic society, has been remedied through the elevation of work's existential importance. As Derek Thompson writes, our modern ideal of work has promised "transcendence and community," but failed to deliver either.

Productivity metrics shape how we evaluate our self-concept, spend our time, and understand our identity. In "Remedying the Malaise of Work," Elliott Jones examines this phenomenon by comparing the works of the religious philosophers Josef Pieper and Abraham Joshua Heschel. These two writers challenge the assumption that human life should be organized primarily around work. Drawing on Pieper's account of leisure as the foundation of culture and Heschel's philosophy of the Sabbath as sacred time, the paper asks whether meaningful human life actually necessitates time and activities that are exclusively governed by productivity. If modern life increasingly measures time by what it produces, can there still be room for forms of time that are not meant to produce anything at all? And if those spaces disappear, what might be lost from human life along with them?

Along with the paper's compelling arguments, what makes it truly notable is its method. The author fuses the works of a Polish Catholic philosopher and an American Jewish rabbi, both shaped by the devastation of World War II, allowing a dialogue that bridges these disparate authors. Pieper and Heschel wrote in the same years, in response to the same world, and arrived at strikingly similar conclusions from separate traditions. That convergence is not a coincidence. At a moment when interfaith dialogue does not move beyond polite agreement, Jones' work draws out the philosophical substance of both traditions and finds that the Jewish Sabbath and the Christian concept of leisure are not only compatible, but in some deep sense, the same remedy for the same disease. For readers overworked, overburdened, or experiencing ennui, this paper embodies this pervasive sentiment.

As students and young professionals entering the workforce, this paper addresses the growing sense of stress and exhaustion that many of us already experience, and that often intensifies as work becomes a larger part of our lives. It effectively highlights how modernity's utilitarian ethos sets unrealistic standards that result in self-reproach. This paper calls on readers to partake in leisure and celebration, regardless of religious belief(s), as a balm for burnout and the societal pressure to work without end. Jones ultimately reminds the reader that pause is not merely permitted but indeed integral; rest is imperative

both for better work and, more profoundly, for the fulfillment of life itself. For those reluctantly immersed in Western achievement society, we hope this paper fosters engagement with the worth of leisure.

ABSTRACT

This research paper is a comparative study of the works of the rabbi and theologian Abraham Joshua Heschel and the Catholic philosopher Josef Pieper. Drawing on each thinker as a prominent voice within their respective religious traditions, I present an interreligious and philosophical study of their responses to modernity found, in particular, in Heschel's *The Sabbath* and Pieper's *Leisure: The Basis of Culture*. Starting with the works of Byung-Chul Han, I build a historical and philosophical framework of "modernity" which, I argue, Heschel and Pieper respond to in their works. Drawing from their religious traditions in the aftermath of World War II, each thinker argues against the reduction of human life to a worker's existence, to the quantitative lens of the economy and time, and the attachment to "civilization" and "total work" that often leads to *acedia*, discontentment, and restlessness. My work contributes to the interreligious dialogue between the Jewish and Christian traditions by amplifying the perspectives of two modern religious thinkers who were never able to engage in conversation.

Homo nascitur ad laborem, et avis ad volatum.

(Job 5:7 Vulgate)

INTRODUCTION

In modern Western culture,[1] there is much angst and disparagement of the value of work. One views work as a self-destructive activity or views work as meaningless in itself. Yet, if work is inevitable, the main question is how can we best relate to work positively and understand its meaning and value today? To answer this, we must determine what ordinarily gives work its value.[2]

Modern society's failure to answer these questions stems from its attempt to turn work into an end in itself and turn leisure into a means for more work. This was inevitable, suggests Byung Chul-Han, because it is more economical for a society to say *You Must Work Now Without End* than a society that says *You Must Rest*.[3] Despite its efficiency and speed, such a society inevitably leads

1 While I consider a Western analysis of culture in this essay, the underlying arguments of the nature of human activity is not a claim about a particular culture but a claim on the nature of activity itself. This claim, no doubt, remains to be disputed in a further essay.

2 Work, as conceived by the authors I consider, is not valued in terms of any value theory of price. Work, rather, is valued based on the type of activity it is. Any activity that is pursued as a means to a further end is "work." Any activity that is pursued as an end in itself, is by definition, not work. This presupposes ontological claims on the nature of activity and the ends to which actions are pursued. Some of these presuppositions will be explored in this paper, but the focus will be on how this definition is expressed by various authors. Note also that I use work and labor interchangeably throughout this paper.

3 Han, *The Burnout Society*, 9-11.

to burnout, anxiety, and despair. Consequently, this society fosters the inability to contemplate, to linger, to endure, to pause, and the inability to be attentive. Thus, this society is incapable of leisure, sabbath, festivity, and worship.

This societal incapability for leisure, sabbath, festivity, and worship inhibits human flourishing. Rest is a biological necessity. Festivity and worship are universal anthropological phenomena conducive to happiness and health. Leisure, which is not simply resting, but constitutes moments of "exaltation," is an action that Abraham Joshua Heschel noted is necessary for flourishing. The inability of modern society to engage in contemplation, leisure, sabbath, festivity, and worship was a major concern of Abraham Joshua Heschel and Josef Pieper in their works. All these activities, including worship, are only possible in a culture which sees them as the proper end of work, not the means to more work. Indeed, this is because there is something distinctive about these activities, namely that they can be pursued as ends in themselves without the reduction to productivity, utility, and toil.

At the heart of both Pieper and Heschel's thought is the belief that "true life begins when concern for survival...ends."[4] If this is true, all activity, and consequently, political life, should be ordered to ends which are beyond survival. All action and all societies tend toward this naturally, as Aristotle writes, for "the first principle of all action is leisure."[5] Furthermore, Aristotle argues in his *Metaphysics* that the progression of the servile arts to the liberal arts is the natural state of affairs for any society.[6] In other words, the fruits of any society are not the things needed to survive (e.g., food, housing, medicine, etc.) but the things that transcend survival (e.g., religion, knowledge, art, etc.). Pursuit of what is good, true, and beautiful is necessary to our flourishing precisely because it goes beyond our mere survival. Modern society would be wise to follow the forgotten adage of Aristotle that the pursuit of political life is not for mere living but "exists for the sake of living well."[7] Therefore, political life should pursue those actions that transcend mere survival. Pieper and Heschel can help us discern what such actions are. In addition, Pieper and Heschel provide us with an interreligious dialogue from the Christian and Jewish traditions as they seek to propose remedies for the ills of modern society. In doing so, their works extend beyond religion and propose questions of philosophical importance on the nature of action and human flourishing.

SECTION I. WHAT IS MODERNITY?

Yet living as we do in an age of superficiality, rushing frenetically from one thing to another without really knowing why, and ending up as insatiable consumers and slaves to the mechanisms of a market unconcerned about the deeper meaning of our lives, all of us need to rediscover the importance of the heart. (Delexit nos, §1)

---

4 Han, *Vita Contemplativa: In Praise of Inactivity*, 3.
5 Aristotle, *Politics*, 1337b23-1338a30.
6 Aristotle, *Metaphysics*, 981b15-25.
7 Aristotle, *Politics*, 1252b.

For our purposes, "modernity" will refer to the historical period encompassing the lives of Josef Pieper (1904-1997) and Abraham Joshua Heschel (1907-1972) and, in particular, to the years after the writings of Pieper's *Leisure the Basis of Culture* (published in 1948) and Heschel's *The Sabbath* (published in 1951). However, I must preface that my account of modernity does not historically follow every moment in that period. Instead, I describe a *zeitgeist*, a mood *(Befindlichkeit)*, and the conditions of belief, rather than a year-by-year analysis of historical happenings. Aspects of this construction are not restricted only to this period, as I refer to ideas that preceded and proceeded from this period. But, as will be demonstrated, my construction of modernity is what Pieper and Heschel are responding to in their works. While this construction of modernity aims to be persuasive for the present argument, it is acknowledged as a limited construction, consistent with the inherent limitations of any historical or philosophical account.

Three main ideas constitute my framework of "modernity" in these years. These ideas stem from the writings of Byung-Chul Han and, as will be demonstrated in sections II-III, also constitute the main ideas that Pieper and Heschel are responding to. I present these ideas from Han not because it is the best or clearest explanation of modernity, but because they construct a modernity that is precisely the modernity that Pieper and Heschel respond to. These three ideas are as follows: I. The Negative Effects of "Achievement Society," II. The Negative Effects of Modern Time, and III. The Negative Effects of Activity.

I. THE NEGATIVE EFFECTS OF "ACHIEVEMENT SOCIETY"

Han, in his book *The Burnout Society*, juxtaposes the prohibitions of "disciplinary society," such as *You Shall Not* or *You Cannot* and the "achievement society" that promotes *You Can*. Precisely because *You Can*, *You Ought. You Can* becomes an imperative that must be lived up to. Achievement society imposes a moral *ought* that goes beyond *can*. Typically, we think that "ought implies can" as a first principle of moral reasoning: that we can only be obligated, commanded, or punished for an action if we are capable of performing said action.[8] This has been inverted in Han's "achievement society" in which "can implies ought".[9] Because I *can* or am told I *can* do X, this implies I ought to do X simply because I can. This provokes problems for morality, but for our purposes, it provokes a problem for human living. Unlimited *can* ignore the fact that our capacities have limits conditioned by an array of causes (physical, psychological, socio-economic, racial, gendered, religious, etc.). Achievement society becomes self-destructive because when I fail to do X, which I am bound to do because of limits that are conditioned beyond my control, the only person

8 "For, to call a person to account, to approve, or disapprove of his conduct, who had no power to do good or ill, is absurd. No axiom of Euclid appears more evidence than this." See Thomas Reid, *Essays on the Active Power of Man*, 48.

9 For this idea, also termed the "Tyranny of Mastery", and its consequences in modern political and moral life, see R.J. Snell *Lost in the Chaos: Immanence, Despair, Hope*, 46-47.

to blame is the one who is commanded: namely, the conditioned subject.

What appears to be freedom, of endless possibility, is actually a tyrannical relationship. I impose on myself a command to which I am bound to fail.[10] This, for Han, not only creates anxiety (because of the need to fulfill the demand) but also eventually burnout (because I fail to fulfill the demand). *Can* without limit leads to endless achievement and endless production.[11] Herein lies the connection to the economy. It is precisely the economy in an achievement society that thrives. Efficiency results from the subject who feels they are free and in control of their actions, identities, etc.[12] However, for the person, failure is inevitable because endless work is the result of an achievement-oriented society. Failure is inevitable because endless work is only compatible if humans were machines or *animal laborans*.[13] In summary, Han claims that our relationship to work and human living is negatively affected when we are commanded to achieve.

II. THE NEGATIVE EFFECTS OF MODERN TIME

Han, in his *The Scent of Time*, proposes that modernity has led to the loss of structured, rhythmic, and meaningful time. Time is no longer leading to a goal such as the eschaton.[14] Instead, time leads into an undefined future of progress precisely because it has discarded what made time meaningful: namely, God. With the "death of God" the only temporal structures are birth and death (we might add the typical aspirations of acquiring a house, family, etc.).[15] If time is a progressive line, and if there is no end or goal in sight, then time progresses into nothingness. If different periods or "points" of time lead nowhere and cannot be connected in any meaningful way towards an end, then time becomes fragmented. Time becomes atomized with nothing attracting or repelling each moment.[16] In a similar way in which the societal restraints upon human action were relieved in achievement society from *You Shall No*t to *You Can*, time was relieved from the constraints of a set meaning with the death of God. But while this is certainly liberating, such that new meanings that were not possible before can structure time, it also has the possibility of atomized time as a result of the loss of meaning or narrative.[17] Time turns into something without a narrative. If progress (whether scientific or economic) itself is the goal of time, then this also implies an imperative of acceleration so as to reach the goal faster.[18] The lack of structure and meaning, accompanied by the perception of the acceleration of time, also affects our actions and relationships.

---

10 This, of course, does not just affect our relation to work and action but our identities and personal meaning. It is in Bauman's *Liquid Modernity* that "the burden of pattern-weaving [meaning-making] and the responsibility for failure… [that falls] primarily on the individual's shoulders." See Bauman, *Liquid Modernity*, 8.
11 Ibid., 9.
12 Ibid., 47; 49.
13 Arendt, *The Human Condition*, 22.
14 Han, *The Scent of Time*, 4.
15 Ibid., 16.
16 Ibid., 18-19.
17 Ibid., 50-51.
18 Ibid., 30.

There is also a temporal imperative of haste in addition to the moral imperative of achievement society. In other words, because *I can do X, I not only ought to do X, I ought to do X now. I ought to do X now* because either a.) the goal known as "progress" must be reached, or because b.) there is no narrative or meaning to moments of atomized time, so I must compensate by always acting. Han claims, our relation to time and meaning is negatively affected when we are commanded to progress and accelerate for progress's sake.

III. THE NEGATIVE EFFECTS OF ACTIVITY

Han, in his *Vita Contemplativa*, proposes that modernity is characterized by the inability to contemplate, to linger, to endure, to pause, and the inability to be attentive. In *The Burnout Society*, he argues that modern society is characterized not by contemplation but by "hyperattention," in which one is forced to be attentive to every stimuli, precisely because of the lack of the ability to be attentive to one thing.[19] True attention, however, is the ability to not be attentive to every stimulus and attend to one thing.[20] Without this ascetic thwarting of certain stimuli, my experience of the world is one of hyperattention in which I have no power not to focus on X, and am thus bombarded by an endless array of stimuli. Attention means exercising the power not to do or attend to X, which means being able to pause, to rest, to look, to see, to contemplate.[21] The opposite of hyperattention is the vita contemplativa.[22] The vita contemplativa is a life consisting of actions that have the quality of "but to no end"–in other words, it is an action done for its own sake.[23] A life without participation in the vita contemplativa "degenerates into hyperactivity" and results in the burnout described in the achievement society.[24]

In *The Scent of Time*, Han argues that atomized time makes it impossible to "linger" on any moment.[25] Nothing holds our attention if there is no structure or meaning to any moment.[26] As hinted at, in a society conditioned by achievement and atomized time, the imperative to act, and act now, is dominant. The act of resting is not done to rest in a moment or in a thing (atomized time prohibits this), but rest is done for the sake of more work.[27] The human is reduced to an *animal laborans*.[28] Because the type of activity that the *vita contemplativa* presupposes requires a sufficient amount of "free" time, this type of action becomes impossible in a world with accelerating atomic time.[29] Han writes eloquently that "life receives its radiance only from inactivity [actions within the

---

19 Han, *The Burnout Society*, 12-13.
20 Ibid., 24.
21 Ibid., 18-21.
22 Han, *The Scent of Time*, 109-111.
23 Han, *Vita Contemplativa: In Praise of Inactivity*, 5.
24 Ibid., 85.
25 Han, *The Scent of Time*, 40-41.
26 Ibid., 18.
27 Ibid., 92; 98.
28 Ibid., 92.
29 *Vita Contemplativa: In Praise of Inactivity*, 10-11.

vita contemplativa]" and when this is lost, all activity is reduced to labor or an action done for the sake of X.[30] For Han, all action is given a structure when it is for the sake of the *vita contemplativa*.[31] It is because of this idea that Han can write that "true life begins when concern for survival…ends."[32] Inactivity[33] or the capacity to pause, to rest, to hesitate, differentiates the human from the machine or the mere survival-minded animal.

Thus, another imperative has been added to the moral imperative of achievement society: *You can*, and to the temporal imperative of modern time: *Now*. Now, there is the imperative to perform a specific type of action: namely, *Work*. In other words, *because I can do X, I not only ought to do X,* and *I not only ought to do X Now*, but *I ought to do Work Now* because doing work is better than being idle. Therefore, our relation to action is negatively affected when we are commanded to act for action's sake, particularly as defined by *Work*. The effects of all three of these ideas are burnout, anxiety, boredom, meaninglessness, and restlessness.

These three ideas and their effects are either explicitly or implicitly responded to in the works of Josef Pieper and Abraham Joshua Heschel. Either experiencing these negative effects themselves or foreseeing their potential harms, Pieper and Heschel believed it to be their task to propose the remedy and liberation from these harms. I argue that Han's analysis of modern work, diagnoses the ailment to which Pieper's concept of leisure and Heschel's account of the Sabbath offer remedies.

SECTION II. JOSEF PIEPER

At no period have the active, that is, the restless, been of more importance. One of the necessary corrections, therefore, which must be undertaken in the character of humanity is to strengthen the contemplative element on a large scale.

––Friedrich Nietzsche, (*Human, All Too Human,* §285)

As a freelance writer and university lecturer, German-born philosopher Josef Pieper sought to revive both an ancient understanding of philosophy and a Christian understanding of moral values in German social and intellectual life. Pieper sought to console a broken Germany through philosophy, which always drew large attendance in the post-war years.[34] In 1948, *Leisure the Basis of Culture* was published. In this work, Pieper explores the ancient meaning of leisure and its relationship to work, the difference between leisure and idleness,

---

30 Ibid., 2.
31 Ibid., 88.
32 Ibid., 3.
33 The way Han uses "Inactivity" is better defined by Pieper as an activity "that does not have a purpose outside itself, that is meaningful in itself, and for that very reason is …[not] useful." (see Pieper, *In Tune with the World*, 8-9).
34 Ibid., 44.

how leisure relates to philosophy, religion, festivity, economics, and the negative effects of the "world of total work," also termed the "totalitarian work state."[35]

Against the idolatry of work, Pieper proposes that leisure is not idleness and, when properly understood, provides the opportunity for contemplation and worship. Boredom and restlessness are our inability to enjoy leisure, and this has drastic effects for philosophy, religion, human nature, and culture. His thesis is as follows: "Culture depends for its very existence on leisure, and leisure, in its turn, is not possible unless it has a durable and consequently living link with the *cultus*, with divine worship."[36] Human living possesses an inherent structure, defined by the relationship between work and leisure. Within leisure, there is an inherent structure, Pieper thinks, with religion. Because these two structures are inherent to human living, they cannot be removed, and if ignored, as Pieper thinks modern society has done, then human nature and the object of human worship is at risk. With this in mind, three main ideas constitute *Leisure the Basis of Culture* and Pieper's supporting works. These three ideas are as follows: I. Ancient versus Modern Anthropology, II. Festivity and Worship, and III. Leisure as Remedy for Modernity.

I. LEISURE AND WORK: ANCIENT VERSUS MODERN ANTHROPOLOGY

Pieper starts his study in *Leisure the Basis of Culture* by contrasting the ancient versus the modern understandings of human nature. Because it is anthropology that influences one's epistemology, ethics, politics, and consequently the horizon for human living, Pieper is concerned with critiquing and revising a harmful modern definition of human nature.

In modern society, Pieper argues that human nature has been reduced to that of a worker.[37] A worker is not a term designating a contingent social status but rather inherent to what a human is. A reduction in human nature is a reduction in the conditions of possibility of human living. This reduction of human living has been reduced to the horizon, which Pieper calls the "world of total work".[38] The world of total work is a society in which the goal and end of every action is work or reducible to work, to fulfill a need or utility. There is no time nor space that cannot be acted on, toiled, or utilized. The world of total work is certainly in keeping with Han's "achievement society" and the overemphasis on activity in modernity. It is only in a world of total work that an achievement society could not only be a possibility but also be the norm for human action.

To counteract this "world," Pieper proposes the existence of actions that cannot be justified nor achieved within the world of total work. They exist relative to but not reducible to the world of total work. Because the world of total work cannot make sense of these actions, Pieper proposes that the world of

---

35 Pieper, *Only the Lover Sings*, 18.
36 Pieper, *Leisure the Basis of Culture*, 15.
37 Ibid., 22-23.
38 Ibid., 20.

total work is not fully exhaustive and therefore, the modern definition of human nature must be expanded.[39] Such actions that transcend the world of total work fall under leisurely actions done for their own sake[40], such as philosophical contemplation, aesthetic contemplation, disciplines of study such as the *artes liberales* (liberal arts), festivity, and worship.[41] Incorporating such actions into a human life expands the definition of human nature beyond "worker." The implication is that this leads to the flourishing of humanity.

II. PHILOSOPHY OF RELIGION: FESTIVITY AND WORSHIP

The previous subsection described the internal disposition needed for an attitude of leisure. Pieper, however, spends much of his thought describing the philosophical importance of these internal dispositions. Compared with the highest values of the world of total work (i.e., activity, toil, and utility), the "soul of leisure" requires an attitude of passivity, a pause in toil, with no reference to utility.[42] Beyond the formal disposition of the "soul of leisure," the matter or content allows one to confront reality and or the fruit of one's work with the attitude of "it is very good" (Gen. 1:31).[43] Knowing and affirming the fruit of our labor and the reality in which we subsist is cause for celebration. Celebration, thus, is central to the soul of leisure.[44]

What is the motive or intention of leisure? Pieper answers that it is celebration. And what is the purpose of celebration? Pieper answers that it is to celebrate. Celebration as the soul of leisure is an end in itself, pursued for no other reason than to celebrate. Celebration consists of the "affirmation of the universe and [one's] experiencing the world in an aspect other than its everyday one."[45] Not only is one engaged in an activity that transcends the workday world, but one also views this world differently through an internal disposition oriented towards a proper end in itself.

It is celebration, as the soul of leisure, that is the origin of Pieper's philosophy of religion. Celebration, Pieper argues, is the root of festivity. A proper festival, Pieper writes, is defined as "performing and showing, in a not everyday fashion, one's already established affirmation of the world and of existence."[46] This demonstrates the convergence of Pieper's existential philosophy and philosophy of religion.

---

39 That the world of total work cannot make sense of an action done for its own sake that cannot be acted on, toiled, or utilized does not mean such an action does not exist. What Pieper expresses so well, is the incommensurability between action in the world of total work and action that transcends that world. This incommensurability for Pieper is a hierarchy of useful actions to useless actions.

40 It is important to note that nothing prevents an action done for its own sake to have an effect as a byproduct that can be utilized, see *Leisure the Basis of Culture*, 40.

41 Ibid., 37-38.

42 Ibid., 49.

43 Pieper, *In Tune with the World*, 47.

44 Pieper, *Leisure the Basis of Culture*, 65.

45 Id.

46 Pieper, *Not Yet the Twilight*, 26.

Festivals[47] are inherently leisurely because they require celebration, and it is through this celebratory nature that they become inherently religious. Pieper offers proof of this both historically and biblically. Put simply, before modern society, days of rest were always for divine worship.[48] Worship is the same thing for time as a temple is for land, which is to say that a time for worship is an architecture of time: a reserved time for non-work.[49] Recall that this closely resembles Han's idea that the sense of and obligations for the Judeo-Christian God provided a structure to time. It follows that "the most festive festival" is divine worship because the most intense celebration is "praise of God".[50]

Celebration is not mere enjoyment in the state of rest but must include the work that was done and the resulting fruit. The toil and the resultant fruit of work must be celebrated together.[51] Work and its fruit are celebrated precisely by sacrificing them during festivals. A festival is a time apart from the workday that is created from the excess fruit of labor, not only materially but existentially as well.[52] There must be material in excess to sacrifice, and there must be an internal disposition, a richness, to feel the need to sacrifice.[53] Sacrifice, as mentioned above with religion, is antithetical to the world of total work and its highest values of activity, toil, and activity.[54]

The heart of Pieper's existential philosophy and philosophy of religion is the idea that leisure makes possible festivals that require celebration and sacrifice, which is a type of joy, love, and praise.[55] Celebration, as mentioned, is an affirmation of reality that is reduced to the idea that "everything that is, is good, and it is good to exist".[56]

Without celebration or the affirmation of all reality (whether with a religious belief or not), leisure is degraded into mere killing time and boredom.[57] Without affirmation, one is left with a negation that Pieper, the ancients, and the medievals called *acedia* or sloth. *Acedia* for the ancients and medievals is the inability to be leisurely: that is to say, the inability to affirm reality, and human nature, as it is. If there is no use or value in leisure, then there is nothing beyond work for which work is done. *Acedia* is a rejection of leisure because it leads to the attitude of "work for work's sake."[58] Leisure, consisting of an affirmation of the fruit of one's work and of reality, is antithetical to *acedia*.

---

47 Festivals of course are not just adorned by religious ritual but also the arts of singing, dancing, music, etc. Art adorns festivals but can also be its own festive activity in so far as it is a celebration of being (see Pieper, *In Tune with the World*, 52; 54). For a nostalgic description of a festival see the beginning of T.S. Eliot's "East Coker" in the *Four Quartets* (see Eliot, T. S. Four Quartets. New York: Harcourt Brace and Company, 1943. 11-12).
48 Pieper, *Leisure the Basis of Culture*, 67.
49 Id.
50 Ibid., 65.
51 Pieper, *In Tune with the World*, 4-5.
52 Existential abundance or richness is a term Pieper uses in contrast to the existential poverty of the proletariat, eg. *In Tune with the World*, 19.
53 Ibid., 18-20.
54 Ibid., 19; cf. *Leisure the Basis of Culture*, 68.
55 Celebration, however, is not exclusive to festivals, but arguably to all leisurely actions.
56 Ibid., 26. This idea, from Thomas Aquinas, is formulated in one instance as, "*omne ens, inquantum est ens, est bonum*" [All being, in so far as being is, it is good], see Thomas Aquinas, *Summa Theologica*, I, q. 5, a. 3, [Translation own].
57 Ibid., 69.
58 Pieper, *Leisure the Basis of Culture*, 43.

Acedia results from existential poverty or "inner impoverishment". The world of total work is rich in material goods, but poor in "existential richness."[59] This means that it is sloth, the inability to enjoy leisure, that breeds the attitude of work for work's sake. If one sees no good in reality and derives no delight in work or the fruit of work, then one ceases to be leisurely and festive. Consequently, a world in which leisure and festivity are impossible is a reduction to the world of total work. In this world, work is done for more work, and there is no delight in it or its fruits.[60]

As a consequence of this reduction of human nature, leisure is viewed by modernity as "laziness, idleness, and sloth."[61] The inability for leisure is the inability to see the goodness of being, to say "it is very good," to celebrate, to be inactive, and to do something useless. In the world of total work, neither leisure, festivity, nor religion can exist properly speaking because they all require an action done for its own sake without reference to activity, toil, and utility.[62] It is the *acedic* person that Han identifies when discussing a society obsessed with activity. Affirmation rather than negation is central to the "soul of leisure" and is necessary to combat the negative effects of *acedia*.

III. MODERNITY: LEISURE AS REMEDY

This last subsection will discuss leisure as a remedy for modernity using Han's thought in Sec. I. Recall that the three main effects of modernity were I. a harmful relationship to work due to the command to achieve, II. a harmful relationship to time due to the loss of structured time, and III. a harmful relationship to action due to the inability for "inactivity."

Regarding the first effect, Han's "achievement society" is a development of Pieper's "world of total work" or "totalitarian work state." Pieper's world of total work restricts the domain of human action to work and, consequently, the possibilities of human nature to a worker. Likewise, Han's achievement society defines all actions in relation to achievement. No action can be viewed or valued apart from its capacity to achieve, and consequently, the possibilities of

---

59 Pieper, *In Tune with the World*, 19; cf. *Leisure the Basis of Culture*, 68.

60 This result is echoed both by Pieper, see *In Tune with the World*, 28 and by R.J. Snell, see *Acedia and Its Discontents: Metaphysical Boredom in an Empire of Desire*, 63; 66. Acedia is not simply laziness or not wanting to work. It is the inability to accept reality and one's self as it is. It is a hatred of life, reality, and work. The result of which might be laziness or frenetic activity. The inability to pause, rest, and reflect are characteristic of the acedic person. For Pieper, pausing and resting consists of an affirmation or celebration *in things* rather than a rest and pause *from things*.

61 Pieper, *Leisure the Basis of Culture*, 43.

62 The so-called Protestant Work Ethic from Max Weber is an interesting counterexample. However, even though the Protestant Work Ethic started as the utilization of work for the purpose of religion (i.e., as a sign of the elect), it increasingly lead to the utilization of religion for the purpose of work leading to the decline in religious values (from asceticism to consumption). See Wisman, Jon D., and Matthew E. Davis. "Degraded Work, Declining Community, Rising Inequality, and the Transformation of the Protestant Ethic in America: 1870-1930." The American Journal of Economics and Sociology 72, no. 5 (2013): 1077-1079. http://www.jstor.org/stable/43818636. Han also notes this connection between capitalism and the Protestant Work Ethic which elevated the *vita activa* (see *Scent of Time*, 88-90). The *vita activa* was elevated as the will of God because it was thought to be the only activity capable of increasing God's glory: "According to the will of God…only activity, not idleness and enjoyment, serves to increase His glory." See Weber, *The Protestant Ethic and the Spirit of Capitalism*, 142.

human nature are reduced to those of an achiever. Leisure, properly speaking, as consisting of actions done for their own sake, is not possible in Han's achievement society because leisure can only be done for some gain measured by achievement. The command of achievement society––*You can*––is antithetical to the soul of leisure as celebration, which has nothing to do with anything one does or achieves but precisely with what one does not do and what one instead receives. To say "it is good" as an utterance is a response to something that is received. Celebration, then, is a passive attitude. And even if there is celebration in the achievement society, the only thing that can be celebrated is a job well done. Recall that the highest values in the world of total work are activity, toil, and utility. A consequence of this, as Pieper writes, is that the worker "mistrust[s] everything that is effortless; he refuses to have anything as a gift."[63] Instead, the worker is satisfied with constant toil, whose "face is marked by strain and tension"[64] until, using Han's phrase, he experiences burnout. It is this capacity for passivity that marks the soul of leisure. Thus, Pieper proposes a receiver rather than an achiever society: a society founded on gifts rather than products.

As to the second effect, Han's atomized time is a result of the loss not only of worship but of leisure time that serves as an architecture of time. Atomized time has two main features: lack of narrative and the imperative to act *now*. While Pieper does not discuss secularization or the "death of God" in the works we have discussed, it is clear that religion is increasingly threatened as an end in itself in the world of total work. As such, the narrative that it provides is threatened by its reduction to activity, toil, and utility.[65] Furthermore, a world of total work benefits from the imperative to act *now*. Delays, traffic, jams, and inefficiency are all threats to the efficiency desired by the economy. Progress is alien to leisure because it presupposes that something is used as a means for a further end. For Pieper, precisely because there is a time that cannot be reduced to activity, toil, and utility, time is preserved as land is for a temple––in other words, leisure provides a temporal structure.[66] During festivity in particular, Pieper argues a new sense of time is received as a "pure gift."[67] It is the fruit of festivity as a time away from normal time that is described as entering into the "realm of Eternity"–a time beyond time.[68] Such an experience, whether religious or not, Pieper argues, is only possible insofar as there is a time apart from work.[69]

Finally, with regard to the third effect, it is clear that every action in the world of work cannot be viewed apart from activity, labor, and utility. As such,

---

63 Pieper, *Leisure the Basis of Culture*, 36.
64 Ibid., 30.
65 In other words, religion in the world of total work increasingly becomes utilized for the ends of the totalitarian work state.
66 Ibid., 67.
67 Pieper, *In Tune with the World*, 39.
68 For a paper devoted purely to the relationship between Pieper and Byung-Chul Han on the concept of festivity see Šokčević, Šimo, and Tihomir Živić. "Byung-Chul Han and Josef Pieper on Festivity: An Attempt to Rehabilitate the Culture of Festivity at the Time of Mere Survival." Bogoslovska Smotra 91, no. 5 (2022): 915–41. https://doi.org/10.53745/bs.91.5.6.
69 Pieper, *In Tune with the World*, 41-43.

every action is reduced to a type of work. There is no *vita contemplativa* in the world of total work. In fact, Pieper writes extensively on the distinction between the *vita activa* and *vita contemplativa*––prioritizing the latter. Pieper uses this distinction to make an analogy between work and leisure: "Leisure, like contemplation, is of a higher order than the *vita activa*."[70] Leisurely actions transcend activity as mere work. Leisurely actions are a passive activity and thus transcend the world of total work. Pieper, as a philosopher, focuses heavily on philosophy as a leisurely activity. It is the philosophical act, he writes, that is "incommensurable with the world of supply and demand."[71] This is directly contrary to Karl Marx, who imagines philosophy as a productive science.[72] Leisurely actions are antithetical to hyperactivity. Leisure presupposes the ability to not focus on X. As Pieper writes, "no one can perceive anything unless he is quiet; only the silent person can hear things."[73] It is Han's hyperactivity that Pieper argues is degrading human nature by not allowing a person to contemplate reality. As Pieper writes, "the average person of our time loses the ability to see because *there is too much to see!*"[74] This is precisely the inability to not focus on X, which Pieper is referring to. Leisure presupposes an attention contrary to hyperactivity and is itself an action unconcerned with any product. A truly flourishing society must therefore pursue leisure as an end in itself lest it fall into the world of total work and the slothful rejection of work and reality. And of course, because Pieper ties leisure to religion, a truly flourishing society includes religious worship and festivals. Thus Pieper introduces a modern argument for religion not based on revelatory claims or even of religious freedom, but rather based on human flourishing and the nature of action.[75]

---

70 Pieper, *Leisure the Basis of Culture*, 50. This distinction in turn is best explained by Aquinas in ST. II-II.182; cf. Aristotle, Nicomachean Ethics 1178b24-33.

71 Ibid., 78-79.

72 "The philosophers have only interpreted the world, in various ways; the point is to change it." See Karl Marx, Theses On Feuerbach, Chapter 11. https://www.marxists.org/archive/marx/works/1845/theses/theses.htm; cf. Pieper, *Leisure the Basis of Culture*, 91.

73 Pieper, *No One Could Have Known*, 65.

74 Pieper, *Only the Lover Sings*, 32; cf. 33. In losing the capacity to "see" Pieper writes "there exists a limits below which human nature itself is threatened, and the very integrity of human existence is directly endangered."

75 I want to note that modern readers might be aghast at the ties of human flourishing, leisure, and religion together. Objections might claim that religious institutions command just like achievement society does. For example, they might command something like, "You Ought to Pray Now." As a result, worship feels more like work than a leisurely activity. These objections, however, do not stand against Pieper's arguments. First, while it is true that religious practices might feel like work, this does not mean that they are by their nature work. In fact, most leisurely actions will *feel* like work not because they *are* work but because they are activities. Again, work was above defined as an action that is pursued as a means to a further end. Thus religious actions are defined not by their feeling or the fact that they are commanded, but by the nature of the activity they are. Secondly, this objection mistakenly believes that a command is essentially related to the nature of work. A command is essential to the nature of achievement society but not to work. That someone is commanded to act does not mean that that action is work. And thirdly, the objection mistakenly believes that religious obligations are commanded imperatives. On the contrary, against popular belief even the Ten Commandments are not written as a Categorical Imperative ("Do Not Kill"), but rather as a hypothetical promise ("If You Love Me, You Will Not Kill"). See André LaCocque's "Thou Shalt Not Kill", 82-88 in Lacocque, André, Paul Ricœur, and David Pellauer. *Thinking Biblically: Exegetical and Hermeneutical Studies*. University of Chicago Press, 1998. This idea will be important to keep in mind for understanding the Sabbath as a religious obligation and the Fourth Commandment.

SECTION III. ABRAHAM JOSHUA HESCHEL

Let's go, my beloved, to meet the bride, Let us welcome the presence of Shabbat. (Lekha Dodi)

Before Heschel was known as a prominent social justice activist in the 1960s and 1970s, Heschel was a trained rabbi and scholar in the Judaic tradition. Heschel sought to revive the decline and secularization of ritual in Jewish communities, such as kosher, prayer, and the Sabbath.[76] His 1951 book *The Sabbath: its Meaning for Modern Man* can be viewed as a response to an American Jewish populace who, in the process of entering into the American middle and upper class, either abandoned or privatized their Jewish tradition and thereby bought into the consumerism of modernity.[77] His goal in *The Sabbath* was to revive the importance of the ritual of the Sabbath for American Jews on the basis not of merely following tradition but of being aware of the presence of the divine.

Three main ideas constitute *The Sabbath* and Heschel's supporting works. These three ideas are as follows: I. Sabbath versus Civilization, II. Wonder and Piety, III. Sabbath as Remedy for Modernity.

I. THE SABBATH VERSUS CIVILIZATION

Heschel starts his study in *The Sabbath* by comparing what he calls "technical civilization" or the "realm of space" and "the Sabbath" or the "realm of time." Technical civilization is a society[78] that presupposes and enforces humans' desire "to subdue and manage the forces of nature."[79] The conquest of nature is a conquest of space that is done at the expense of time. The conquest of nature is done to get products. Products are not bad in themselves, but Heschel argues that we cannot be enslaved to things and therefore to a society dominated by controlling and producing things in space.[80] It is Judaism's emphasis on time that Heschel proposes is a solution to the dangers of technical civilization. In particular, Heschel focuses on time as central to festivals and religious commemorations of events. Among such events is the Sabbath. The Sabbath provides an "architecture of time" that structures time so as to prevent a perpetual work week. The Sabbath in particular, because of its quality of holiness as declared by God (Gen. 2:3), is a time not just for rest but for turning to "what is eternal in time, to turn from the results of creation to the mystery of creation; from the world of creation to the creation of the world."[81] In other words, it is a time specifically dedicated to religious experience, expression, and other leisurely activities.

---

76 Ibid., 40; 48-51. For statistics on the decline of religious practice in Jewish communities in the 20s and 30s, see Ibid., 51; cf. 215 for a recent 2013 *Pew Research Center* study of Jewish communities.
77 Ibid., 78; cf. Heschel, *The Sabbath*, xii.
78 See note 5.
79 Heschel, *The Sabbath*, 4.
80 Ibid., 6.
81 Ibid., 10.

The Sabbath, as an architecture or "cathedral" in time, provides a structure of action apart from technical civilization.[82] Work is a means to the end of the Sabbath, not for the purpose of resting so as to be more efficient for more work.[83] The problem with technical civilization for Heschel is that the desire to dominate nature becomes an obsession, and we fall victim to consumerism and the covetousness of things. Despite the harsh critiques of technical civilization, Heschel recognizes the dignity involved in labor and its fruits. Advocating or elevating the Sabbath does not belittle labor but instead provides it meaning. Without the Sabbath, all labor risks reducing to mere and endless toil.[84] It is independence rather than renunciation of technical civilization that Heschel proposes. As he writes, during the Sabbath, "we abstain primarily from any activity that aims at remaking or reshaping the things of space."[85] The Sabbath presupposes an independence of utility, things, and the making and producing of money and goods.

The Sabbath is a solution to Heschel's pronouncement of the effects of technical civilization that "labor without dignity is the cause of misery; rest without spirit the source of depravity."[86] The Sabbath is not mere rest from work but presupposes an "existential richness," as Pieper puts it. The solution the Sabbath provides is *Menuha*, which is an inner and external "tranquility, serenity, peace, and repose."[87] It is a type of happiness, repose of mind, and state of the soul. If there is reason to be happy, then there is reason to celebrate. *Menuha* is "a restfulness that is also a celebration."[88] Herein lies one of the connections between Sabbath and celebration. As Pieper reminds us, celebration is a type of praise, and the most intense celebration is the praise of God. It is Heschel who remarks that "The Sabbath teaches all beings whom to praise."[89] Observing the Sabbath is beyond mere rule-following; instead, the Sabbath is a time to "celebrate the creation of the world…"[90] It is in this way that Jews imitate God, who rested on the seventh day. It is celebration that is central to Heschel's reflections on the Sabbath in contrast to technical civilization.

Human nature in modernity is just as important to Heschel as it is to Pieper. As Heschel writes, "what we are depends on what the Sabbath is to us."[91] In other words, who we are, our possibilities of human living, our existential richness or poverty, and our happiness depend on our relation to the Sabbath. Heschel calls this "inner liberty." The Sabbath provides an "inner liberty" from the "domination of things as well as from domination of people."[92] This inner liberty can be understood as providing us not only with a detachment of things

---

82 Ibid., 8.
83 Ibid., 14.
84 Ibid., 28.
85 Ibid., 28-29.
86 Ibid., 18.
87 Ibid., 23.
88 Ibid., xiv.
89 Ibid., 24.
90 Ibid., 19-20.
91 Ibid., 89.
92 Ibid., 89.

but also a changed way of seeing things. We see things and people not just as objects that can serve our interests–interests measured by technical civilization–but instead as objects and subjects capable of serving another end. It is a proper relation to things and people that the Sabbath provides.

The Sabbath presupposes that work and its fruit are not done for their own sake but for the sake of something else: namely, the Sabbath. Heschel describes the days during the week as insignificant in themselves and only have meaning in reference to the seventh day. In the view of technical civilization, the days during the week have significance only in relation to the number of hours worked, profit gained, measured against losses, and progress achieved. The significance of the Sabbath, however, is measured by none of these metrics nor any metric at all, in part because no work is done and because its significance is not achieved from something outside it. The Sabbath is a good in itself, to be enjoyed for no other reason than that it is good and is also holy (*qadosh*).[93] It is the Sabbath that demonstrates that not all times are equal, as "there is a hierarchy of moments within time."[94] Heschel writes that Jews can "work with things of space but… [must] be in love with eternity."[95] Eternity, for Heschel, is the presence of the divine in this world. Eternity is an experience of time, such as during the Sabbath, that renders human consciousness receptive to God. A time for time is really a time for God or the awareness of God. Pieper also remarks in his reflections on a new sense of time during festivals that eternity is implicit in the concept of time and creation.[96]

Sabbath as a concept, as a day, and as consisting of presuppositions is juxtaposed to technical civilization. The Sabbath surpasses the presuppositions and values of technical civilization in a similar way that Pieper's leisure transcends the world of total work. While Pieper's concept of leisure is open to religion and has a religious context, the Sabbath is distinctly religious and rooted in the Judaic tradition. Heschel's pronouncement that "we must conquer space in order to sanctify time"[97] is both recalling God's reminder that "six days you shall labor and do all your work. But the seventh day is a sabbath to the Lord your God" (Ex. 20:9-10) and God's sanctification of the seventh day (Gen. 2:3). It remains to be explored how God is witnessed on the Sabbath and what the effects of this are.

II. PHILOSOPHY OF RELIGION: WONDER AND PIETY

I have already mentioned that Heschel's concept of *Menuha* is similar to Pieper's "soul of leisure." This however, compares Heschel's explicitly Jewish concept with Pieper's irreligious (albeit theological) concept of celebration and affirmation of reality. I want to note a source in Heschel's writings that are less overtly Jewish and more open to philosophical discourse.

---

93 Ibid., 9.
94 Ibid., 96.
95 Ibid., 48.
96 See note 103.
97 Ibid., 101.

Heschel writes in his 1951 book *Man is Not Alone: A Philosophy of Religion* that there are two faculties through which we apprehend the world: namely, "with reason and with wonder."[98] The latter is a state of the mind in which "we are amazed at seeing anything at all; amazed…at the unexpectedness of being as such, at the fact that there is being at all."[99] While philosophy starts with asking the question of *why is there anything at all rather than nothing*, wonder is the state that allows one to ask this question. This sense of wonder, Heschel calls "radical amazement," and the object of wonder is the "ineffable."[100] If for Pieper all being is good in so far as it is, for Heschel all being is ineffable is so far as it is. The experience of the ineffable is one that alludes all concepts of reason.[101] A typical response that accompanies this awareness of reality is "awe or reverence."[102] Awe at the meaning of things transcends conceptual formation and the common effect is to receive, "to pause, to be still, [so] that the moment may last" rather than trying to grasp, achieve, or skip the moment.[103]

The distinction between reason and wonder can also be likened to Heschel's distinction between technical civilization and the Sabbath. The reason that dominates technical civilization is a reason concerned with "expediency," accumulating information or utilizing reason so as to dominate nature, expand power, etc. The world under this rational attention "turns out to be alien and weird."[104] The wonder that gives birth to the arts, philosophy, poetry, and religion (and consequently the Sabbath) on the other hand is an attention concerned with "appreciation." The world under this wonderful attention turns out to be amazing, awe-inspiring, and meaningful. It is in moments of radical amazement, then–and the interpretation of them as absolutely meaningful–that can be a remedy for existential poverty and the misery of the worker without rest or spirit. It is reason's inability to grasp the ineffable in moments of radical amazement that gives birth to the arts such as music, poetry, and religion.[105]

However, just like the inability for true leisure to exist in the world of total work, Heschel writes that it is technical civilization that threatens "the sense of wonder."[106] Heschel's concern with technical civilization is a concern regarding human happiness and fulfillment: "The beginning of our happiness lies in the understanding that life without wonder is not worth living."[107] When humans are aware of the ineffable, they are no longer separated from the world under

---

98 Heschel, Man is Not Alone: A Philosophy of Religion, 11.
99 Ibid., 12.
100 Ibid., 13. It is important to note that this sensus ineffabilis is an *a priori* postulate of the mind similar to John Calvin's *sensus divinitatis*, see Ibid., 16-17; cf. 20-21.
101 To grasp this idea it is best to consider an experience which produces many ideas but no determinate totalizing concept, e.g., the face of another, one's body, an unexpected event, an artwork or icon. This idea of Heschel's is close to Immanuel Kant's description of aesthetic ideas in the *Critique of Judgement*: "…by an aesthetic idea I mean a presentation of the imagination which prompts much thought, but to which no determinate thought whatsoever, i.e., no [determinate] concept, can be adequate…" Kant, *Critique of Judgment*, 182.
102 Heschel, Man is Not Alone: A Philosophy of Religion, 23.
103 Ibid., 26.
104 Ibid., 37.
105 Ibid., 36.
106 Ibid., 37.
107 Id.

the attention of "expediency" or technical civilization, which views all reality as things to be used for man's needs. It is wonder or radical amazement at the ineffable that views all reality as full of meaning.

Heschel proposes that there are two further responses to the ineffable that. One is praise,[108] which is similar to Pieper's concept of celebration, and the second is the moral awareness of being needed by another.[109] This moral awareness is both out of devotion to the ineffable, or God, and to other humans. It calls us to action, to change ourselves and to support the lives of others around us. Heschel does not fully flesh out how this moral awareness is achieved but if it could be formulated it would be something similar to Franz Rosenzweig's description of God's commandment: " Love Me!"[110] Religion begins then with a sense of something asked of us following an awareness of the ineffable.[111]

What is implicit in Heschel's philosophy of religion is that the Sabbath is a type of action in which an awareness of the ineffable is possible and, consequently, wonder, awe, reverence, praise, and an ethical awareness towards God and others. The Sabbath, rather than technical civilization, is an answer to the "urgencies, the perpetual emergencies of human existence, [and] the rare cravings of the spirit."[112] It is wonder and awe at reality and the practice of the Sabbath that is a remedy for the ailments of technical civilization.

III. MODERNITY: SABBATH AS REMEDY

This last subsection will discuss the Sabbath as a remedy for modernity using Han's thought in Sec. I. Recall once again, that the three main effects of modernity were I. a harmful relationship to work due to the command to achieve, II. a harmful relationship to time due to the loss of structured time, and III. a harmful relationship to action due to the inability for "inactivity".

Regarding the first effect, it is clear that Heschel's Sabbath is antithetical to Han's "achievement society." Han's achievement society defines all actions in relation to achievement. It is the Sabbath that defines action not in relation to any utilitarian value but is instead valued for its own sake because it is a holy good given by God: "the hours of the seventh day are significant in themselves."[113] The command of achievement society––*You Can*––is antithetical to the Sabbath, which does not propose endless action for the sake of achievement or the standards of civilization but instead abstains from said actions and from the things of space. This, however, does not imply that during the work week a Jew adopts the commands of "achievement society;" on the contrary, all action during the work week is done for the sake of the Sabbath.

The concept and inner attitude of *Menuha* is contrary to the command to achieve. It is *Menuha* that allows one to rest in the good in celebratory praise. It is hard to conceive the achiever in an achievement society not only resting but also

108 Ibid., 74.
109 Ibid., 69.
110 Franz Rosenzweig, *Star of Redemption*, 192.
111 Heschel, Man is Not Alone: A Philosophy of Religion, 215
112 Heschel, God in Search of Man: A Philosophy of Judaism, 372.
113 Heschel, *The Sabbath*, 20.

resting in an intrinsic good instead of striving for another good. In fact, Heschel writes that Judaism and by extension religion, "plants in man…a need of spiritual needs rather than a need of achievements, teaching him to be content with what he has, but never with what he is."[114]

As to the second effect, Heschel's reflections on time and the Sabbath are contrary to the negative effects of modern time. It is the Sabbath that provides an architecture of time, solving the problem of atomized time and the lack of narrative. The Sabbath provides a structure in time to a week, but the Sabbath itself is both a remembrance and a hope for a future. God commands the Israelites to "remember the Sabbath day" (Ex. 20:8-11). This act of remembrance recalls God's rest from creation, and in doing so, serves as a call to "imitate the creator, to be like God."[115] The Sabbath also provides a structure to the future through hope. While Heschel does not go into detail about the Jewish concept of the afterlife, he does remark that the Sabbath is a taste of the "world to come."[116] Above all, the Sabbath opens one to the experience of eternity in time.[117] Regarding the imperative to act *now*, there is no such hastiness in the Sabbath. The Sabbath is welcomed and waited for and given rather than reached by man's effort. Heschel argues that the negative command of the tenth commandment of "You shall not covet," can be paradoxically interpreted as an imperative to "covet things of time."[118] Despite this imperative, it is clear that coveting time is not something that can be reached by speed, efficiency, and urgency but instead requires waiting and exciting expectation.[119]

Finally, as to the third effect, it is clear that the Sabbath is antithetical to a society that reduces all action to desiring things of space. Technical civilization tries to satisfy our own desires rather than God's, which is why it is susceptible to reducing rituals, especially the Sabbath, to some utilitarian benefit like health or happiness.[120] Heschel astutely remarks that "He who performs a ritual does not expect to advance his interests…The Mitzvot is a form of expressing in deeds the appreciation of the ineffable."[121] The Mitzvot, therefore, is a way of expressing praise for the ineffable and is detached from any utilitarian benefit they might produce.

As argued above, the Sabbath opens one to an experience of the ineffable, in which it is possible to be aware of a command to action. However, this moral command, a sense of *you ought*, is measured not by human needs but by God's needs or commands. While Heschel advocates for action inspired by the moral imperative, it is measured by an altogether different standard than civilization.[122]

---

114 Heschel, *Man is Not Alone*, 257. Of course, spiritual needs require striving but this is detached from the metrics of civilization.
115 R.J. Snell, *Acedia and its Discontents: Metaphysical Boredom in an Empire of Desire*, 103.
116 Heschel, *The Sabbath*, 19; 73-74.
117 Ibid., 75-76.
118 Ibid., 91.
119 Han himself writes that there is in the Sabbath a suspension of ordinary time in "blissful inactivity." (See Han, *Vita Contemplativa: In Praise of Inactivity*, 60).
120 Heschel, *God in Search of Man: A Philosophy of Judaism*, 350.
121 Id.
122 See for example Zelizer who writes: "The heart of Heschel's philosophy of religion and of Judaism is piety consisting of "adher[ing] to tradition, to open oneself to God, and to devote one's entire being to pursuing social justice in the hear and now." See Zelizer, *Abraham Joshua Heschel: A Life of Radical Amazement*, 224.

Regardless of the emphasis on action in this aspect of Heschel's thought, it is clear that all actions during the Sabbath are detached from any metric of civilization's value. This consequently enables the possibility of actions done for their own sake that are considered inactive from the standpoint of civilization. The Sabbath is a time consisting of actions irreducible to those of work, and this is directly contrary to the negative effects of modern action. For Heschel, the Sabbath is so integral to society, that he wrote the future of civilization "will depend upon how much of the Sabbath will penetrate its spirit."[123] The Sabbath's spirit is detached from achievement, the economy, utilitarian gain, atomized time, urgency, hastiness, and lastly, work. It is Judaism and the practice of the Sabbath that is a remedy for the ailments of technical civilization.

SECTION IV. COMPARATIVE REFLECTIONS

But the Lord answered her, "Martha, Martha, you are worried and distracted by many things; there is need of only one thing. Mary has chosen the better part, which will not be taken away from her."––Luke 10:38-42 (NRSVCE)

The negative effects of modernity, as described by Han, Pieper, and Heschel can be broadly summarized as sloth and gloom.[124] It is the slothful and gloomful person that succumbs to the temptation that *I ought to do Work Now.* Why? Succumbing to the imperative that I ought to work for work's sake fundamentally disrupts a natural hierarchy of action, reducing human nature to the sphere of total work and a world of technical civilization concerned with immediate human needs and the things of space. Neither action nor thing can transcend this world. There is nothing good in itself, and there is nothing holy. There is no celebration, festivity, or religion properly speaking. The effects of all three of these ideas are burnout, anxiety, boredom, meaninglessness, and restlessness. Leisure is a response to sloth, and the Sabbath is a response to gloom. As R.J. Snell reminds us, to counteract sloth requires "cultural forms and practices, not merely lectures and ideas," emphasizing the need for a transformative change in disposition.[125] Both Pieper and Heschel provide not only ideas but practices of leisure and Sabbath as remedies.

Regarding their religious influences, while leisure itself is not exclusively Christian, the Sabbath is uniquely Jewish. While Pieper is more indebted to ancient Greek and medieval philosophy in his works, his conclusions never

---

123 Heschel, *God in Search of Man: A Philosophy of Judaism*, 418.
124 The gloomful person is antithetical to the pious person for Heschel, see *Man is Not Alone*, 286-288.
125 R.J. Snell, *Acedia and its Discontents: Metaphysical Boredom in an Empire of Desire*, 97.

veer far from a Catholic understanding.[126] For example, Pieper's fears of the reduction of action to utility are mirrored by Catholic social teaching in the Catechism[127] and many encyclicals such as *Laborem exercens*,[128] *Centessimus annus, Gaudium et spes*, etc. Both Pieper and Heschel argue for a hierarchy of action in which work is done for leisure and work is done for the sake of the Sabbath.[129] It must be noted that further exploration could be dedicated to the religious motivations and intellectual influences that underpin their respective works.

Lastly, it is important to note that ecumenical dialogue was in keeping with the beliefs and lives of both Pieper and Heschel. Pieper himself was involved in many Christian ecumenical groups and conferences prior to Vatican II.[130] Heschel was involved in ecumenical groups during the civil rights movement and as a representative of the Jewish delegation to Vatican II, where he advocated for revised language concerning Jews in the encyclical *Nostra aetate*.[131] This important aspect of Pieper's and Heschel's legacy has been deliberately excluded from this paper to maintain a focused philosophical study of their works.

What Pieper's and Heschel's works effectively argue is that "only a people instructed or trained to see and celebrate abundant richness can approach the rest of their lives, the other six days and the work of those days, with a festive eye."[132] Recall that festivity is a type of leisurely action (which is also central to the Sabbath[133]), consisting of an affirmation of reality in the form of a celebration.

There is one notable source of disagreement between Heschel and Pieper. First, it is clear that for Heschel the awe and wonder at the goodness of reality does not stop at aesthetic appreciation. It extends beyond aesthetic appreciation to gratitude to God as well as a moral concern for other people. This ethical dimension, characteristic of many post-World War II philosophers, is noticeably

---

126 See for example Augustine who in his *City of God* uses leisure, Sabbath, festivity, and celebration as themes to describe beatitude. Beatitude will consist in a Sabbath of "eternal rest": "We ourselves shall become that seventh day…There we shall have leisure to be still, and we shall see that he is God…There we shall be still and see; we shall see and we shall love; we shall love and we shall praise." (See *City of God*, 22.30, 1090-1091). See also Augustine in his *The Literal Meaning of Genesis* who joins both Pieper's reflections on leisure and Heschel's reflections on the Sabbath in a meditation on creation: "Finally, he himself neither grew tired when he created things, nor found it a relief when he stopped; but he wished through his scripture to urge upon us a desire for rest by intimating to us that he had sanctified the day on which he himself rested from all his works. At no time, I mean, during the six days in which all things were created is he said to have sanctified anything… but he wished to sanctify this day, on which he rested from all his works that he had made, as though with him also, who experiences not the slightest toil or trouble in his work, rest and quiet is worth more than any activity. The gospel indeed makes the same point for us in the case of human beings, where our savior says that because Mary, seated at his feet, was resting quietly in his word, hers was the better part rather than Martha's, although she out of the loyal devotion with which she was attending to his needs was busy with many things, and working what was indeed a good work." (See *On Genesis*, 4.14, 255-266).

127 E.g., *The Catechism of the Catholic Church*, §2423; §2424; §2426-§2436.

128 E.g.: "Man ought to imitate God both in working and also in resting, since God himself wished to present his own creative activity under the form of work and rest." See LE, §25.

129 Both of these views is held by the Catholic church in *Gaudium et spes*: "Hence, the social order and its development must invariably work to the benefit of the human person if the disposition of affairs is to be subordinate to the personal realm and not contrariwise, as the Lord indicated when He said that the Sabbath was made for man, and not man for the Sabbath." See GS, §26.

130 Pieper, *Not Yet the Twilight*, 24-25.

131 Zelizer, Abraham Joshua Heschel: *A Life of Radical Amazement*, 142-149; 162-163.

132 R.J. Snell, *Acedia and its Discontents: Metaphysical Boredom in an Empire of Desire*, 107.

133 See notes 138; 140-142.

absent from Pieper's writings on leisure and the revitalization of culture. While the actions he proposes are resistant to the world of total work, they are more about revitalizing cultural and political norms rather than ethical encounters. Justice is not something that Pieper sees immediately tied to leisure. While Heschel, an ethical call is a subsequent response to the encounter with the ineffable. Thus, he can more easily integrate ethical obligations to other people as an expression of the Sabbath's spirit. On this note, Heschel's corpus as a whole is more in line with ethically concerned philosophers such as Martin Buber and Emmanuel Levinas who were all concerned with the ethics of face-to-face encounters.

While Pieper on the other hand is not concerned (at least in these writings) with the ethical encounter or moral obligations.[134] I will say however, that the type of ethical actions that Heschel might propose (e.g., actions that fall under the ethics of hospitality such as care for the widow, the orphan, the migrant, and the poor) are fully compatible with Pieper's notion of leisure and actions as ends in themselves. Indeed, Pieper missed the opportunity at building a strong argument against utilitarian ethics and policies that reduce people to metrics and means to ends rather than as values and ends in themselves.

Beyond their convergences and major divergences I lastly want to note the philosophical importance of their works. They raise important philosophical questions within metaethics as to the nature of action. For example, what does it mean for an action to be an end in itself? How do we know what these actions are? How do we pursue them? Which actions are means to further ends? How do we pursue these actions? What are the effects of both of these types of actions? Which actions are necessary to survival and which are necessary to flourishing? Is there a hierarchy of actions and does this lead to the transcendent? These questions are all philosophically important within metaethics and deserve a much further discussion. I must commend Heschel and Pieper who have so convincingly made religious and political use of these philosophical ideas in response to the ailments of modern civilization.

CONCLUSION

This study has compared the writings of Josef Pieper and Abraham Joshua Heschel and their responses to modernity–a modernity that I have described through the works of Byung-Chul Han. I have argued that my construction of modernity is precisely the modernity that Pieper and Heschel are responding to in their works. In particular, Han's ideas of "achievement society," modern time, and modern activity are all seen in various aspects in Pieper's and Heschel's writings.

This study contributes significantly to the field of Christian-Jewish studies and interreligious dialogue. In particular, I have studied their religiously

---

134 Pieper, of course, did write on the cardinal and theological virtues during the 1930s to counter the rise of the Nazi party, see for example *Faith, Hope and Love* (1998), Cf. *No One Could Have Known*, 98.

informed philosophy, which is both a philosophy in response to modernity and a philosophy of religion. A comprehensive comparative study of the major works of Pieper and Heschel has not been extensively undertaken in this manner, and this paper aims to contribute to the long-lasting dialogue between Christian and Jewish thinkers. Lastly, it proposes further questions in the nature of action that extend beyond religious studies into metaethics and human flourishing.

BIBLIOGRAPHY

Arendt, Hannah, and Margaret Canovan. *The Human Condition*. Chicago: Univ. of Chicago Press, 2010.

Aristotle, and Carnes Lord. *Aristotle's Politics*. The University of Chicago Press, 2013.

Aristotle. *Metaphysics*. Translated by Hippocrates G. Apostle. Merrimack, New Hampshire: Thomas More College Press, 2021.

Augustine. *Concerning the City of God against the Pagans*. Translated by Henry Bettenson. Penguin Classics. London, England: Penguin Books, 1984.

Augustine. *On Genesis*. Edited by Boniface Ramsey. Brooklyn, N.Y.: New City Press, 1990.

Aquinas, Thomas. *Summa Theologica*. https://www.thelatinlibrary.com/aquinas/q1.5.shtml. Bauman, Zygmunt. Liquid Modernity. Cambridge: Polity Press, 2000.

Friedman, Maurice S. *Abraham Joshua Heschel: Philosopher of Wonder ; Our Thirty-Year Friendship and Dialogue*. Eugene, Or: Cascade Books, 2012.

Han, Byung-Chul. *The Burnout Society*. Translated by Erik Butler. Stanford, California: Stanford Briefs, an imprint of Stanford University Press, 2015.

Han, Byung-Chul. *The Scent of Time: A Philosophical Essay on the Art of Lingering*. Translated by Daniel Steuer. Reprinted. Cambridge, UK Medford, MA: Polity, 2020.

Han, Byung-Chul. *Vita Contemplativa: In Praise of Inactivity*. Translated by Daniel Steuer. Medford: Polity Press, 2024.

Heschel, Abraham Joshua, Ilya Schor, and Susannah Heschel. *The Sabbath: Its Meaning for Modern Man*. Paperback edition. New York, NY: Farrer, Straus and Giroux, 2005.

Heschel, Abraham Joshua. *God in Search of Man: A Philosophy of Judaism*. New York: Farrar, Straus and Giroux, 1955.

Heschel, Abraham Joshua. *Man Is Not Alone: A Philosophy of Religion*. New York: Farrar, Straus and Giroux, 2006.

Kant, Immanuel, and Werner S. Pluhar. *Critique of Judgment*. Hackett Publishers. 1987.

Lacocque, André, Paul Ricœur, and David Pellauer. *Thinking Biblically: Exegetical and Hermeneutical Studies*. University of Chicago Press, 1998.

Pieper, Josef. *A Journey to Point Omega: Autobiography from 1964*. Translated by Daniel J. Farrelly. South Bend, Indiana: Saint Augustine's Press, 2020.

Pieper, Josef. *In Tune with the World: A Theory of Festivity*. South Bend, Ind: St. Augustine's Press, 1999.

Pieper, Josef. Leisure: *The Basis of Culture ; The Philosophical Act*. San Francisco: Ignatius Press, 2009.

Pieper, Josef. *No One Could Have Known: An Autobiography, the Early Years, 1904-1945*. San Francisco: Ignatius Press, 1987.

Pieper, Josef. *Not Yet the Twilight: An Autobiography 1945-1964*. Translated by Daniel J. Farrelly and Una Farrelly. South Bend, Indiana: St. Augustine's Press, 2017.

Pieper, Josef. *Only the Lover Sings: Art and Contemplation*. San Francisco: Ignatius Press, 1990.

Reid, Thomas. Edited by Baruch A. Brody. *Essays on the Active Powers of the Human Mind*. Charlestown Mass., 1815. Cambridge, Mass: M.I.T. Press, 1969.

Rosenzweig, Franz, and Barbara E. Galli, eds. *The Star of Redemption. Modern Jewish Philosophy and Religion*. Translations and Critical Studies. University of Wisconsin Press, 2005.

Snell, R. J. *Acedia and Its Discontents: Metaphysical Boredom in an Empire of Desire*. Chicago: Angelico Press, 2016.

Snell, R. J. *Lost in the Chaos: Immanence, Despair, Hope*. Brooklyn, NY: Angelico Press, 2023.

Weber, Max. *The Protestant Ethic and the Spirit of Capitalism with Other Writings on the Rise of the West*. Translated by Stephen Kalberg. 4th ed. New York: Oxford University Press, 2009.

Wisman, Jon D., and Matthew E. Davis. "Degraded Work, Declining Community, Rising Inequality, and the Transformation of the Protestant Ethic in America: 1870-1930." *The American Journal of Economics and Sociology* 72, no. 5 (2013): 1075–105. http://www.jstor.org/stable/43818636.

Zelizer, Julian E. *Abraham Joshua Heschel: A Life of Radical Amazement*. Jewish Lives. New Haven: Yale University Press, 2021.

Elliott Jones is a senior studying Philosophy at Boston College. After graduating, Elliott will pursue an MA in Philosophy. His Senior Honors Thesis explores Baroque Scholastic influences in Descartes's theory of ideas. Elliott is interested in the intersection of philosophy of religion in early modern and contemporary philosophy, natural law debates in political philosophy and ethics, and the phenomenology of value. In his spare time he enjoys hiking and critiquing films.

AUTHOR'S NOTE

This project would not have been possible without the support of the Center for Christian-Jewish Learning at Boston College, who sponsored the research, and the Winston Center for Leadership and Ethics, whose continual encouragement has made the presentation of my research possible. I owe a particular debt of gratitude to Prof. Brian Robinette, who first introduced me to

Abraham Joshua Heschel and suggested his potential connection to Josef Pieper. I am also grateful to Hans Gustafson and Jeannine Hill Fletcher, who graciously vouched for my work and made it possible for me to present this research at the American Academy of Religion Annual Conference. I am also grateful for the stimulating conversations on the relation of command and leisure I had at the Eighth International Conference on Philosophy and Meaning in Life.

A common question I get asked about my research is if I have become more leisurely since completing this project. The answer is yes, and I also note that I conducted much of this research during a spring break spent at a Benedictine monastery and a silent retreat at the beginning of the semester. The silence, the prayer, the nature, the solitude are all things that I cherished in a spirit of leisure.

I had originally planned to include a case study drawn from my experience on the Camino de Santiago as a practice of leisurely activity in the modern world. As happens, I got stuck in the theory and never made it out to the potential resources for praxis. I want to leave the reader with three practical reflections. I) Cultivate praise, love, joy, and sacrifice. To do these requires recognizing the inherent goodness and ineffableness of reality. As a result one is able to delight in one's work and its fruits. II) Cultivate the liberal arts. A sign of a leisurely culture is the appreciation rather than devaluation of the liberal arts. They are one of the clearest examples of activities worthy for their own sake. III) Cultivate the gift. Learning to receive gifts implies passivity rather than activity. There is no burnout in receiving but only in achieving. Only in a horizon of the gift does an "achievement society" lose its hold.

I am grateful to the editorial team at LOGOS for their careful reading and critical comments, which have sharpened this paper considerably. Lastly, a special thanks to Wyatt Sell for introducing me to the wonderful work of LOGOS and encouraging me to submit my manuscript.

Aristotle's Natural Slavery in Politics Reconsidered Through De Anima and Nicomachean Ethics

Abigail So
Cornell University

FOREWORD

Editors: Ella-Rose Cespedes, Louis Arnoult Costafreda, Jackie Dorland, Maria Yaniris Gil, Isabelle Schwartz, Wyatt Sell

As long as slavery has been an institution, many great thinkers have attempted to rationalize some form of it. Aristotle is no exception. His infamous theory of 'natural' slavery—where some people, particularly foreigners of ancient Athens, lack the deliberative capacity in their souls—argues that these 'natural slaves' are better off ruled by masters who do have deliberative capacity. For Aristotle, such a hierarchical arrangement is only just.

Unsurprisingly, this argument has been the subject of much criticism by scholars, with them refuting his notion of natural slavery via moral arguments, and explaining how such an august philosopher could have come to such a conclusion by considering his historical context. But rather than employing external theory to challenge Aristotle, Abigail So has taken after Socrates, showing how, in asserting the truth of natural slavery, Aristotle has contradicted himself.

Abigail So starts by introducing Aristotle's account of natural slavery in *Politics* and analyzing how he goes about identifying if a given person's soul lacks deliberative capacity. She proceeds to combine Aristotle's theory of the soul (different orders of potentiality and actuality, form should follow function) in *De Anima* with the habituation of rational capacities in *Nicomachean Ethics*, to show how the limiting power of tyrannical rule or slavery itself prevents us from determining an enslaved person's potential to develop rational capacity, and thus whether they're a 'natural slave'. So makes an astute observation: within the structure of Aristotle's own philosophy, there is no reliable method of distinguishing between the deliberative capacities of natural slaves and that of free persons based on observed behavior. So concludes by theorizing why Aristotle would hold this view: to maintain the hierarchical structure of the household throughout all levels of society, from collective government to individual soul, an approach he perhaps borrowed from his teacher, Plato.

For Logos, this paper was a clear choice. The argument is elegant, exceptionally clear yet profoundly nuanced, written in accessible yet sophisticated prose. Abigail So has managed a fresh perspective on an ancient problem that still honors Aristotle's philosophy and demonstrates impressive knowledge of both Aristotle's works and the extensive scholarship surrounding this particular issue. This paper, as strikingly simple and profound as it is, exemplifies why not everything is always as it seems.

What Aristotle forgets and what So points out is best put in the words of James Baldwin: "I am what time, circumstance, history, have made of me, certainly, but I am, also, much more than that. So are we all."

ABSTRACT

Aristotle's theory of natural slavery in *Politics* has long generated controversy, due to its ethical implications and its apparent tension with the rest of his philosophical corpus. While much of the scholarship has focused on condemning Aristotle's position on moral grounds or explaining his stance through historical and cultural contextualization, less attention has been paid to the foundational assumptions on which his argument is premised. This paper argues that Aristotle's account of natural slavery is theoretically incomplete because it relies on the implicit assumption that observable behavior reliably reveals deliberative capacity. By reading *Politics* alongside *De Anima* and *Nicomachean Ethics*, I focus on Aristotle's claims about the relationship between the soul and the body, the distinction between potentiality and actuality, and the role of habituation in the development of rational capacities. I show that these commitments undermine Aristotle's attempt to identify natural slaves through bodily characteristics or political behavior under tyranny. Ultimately, I suggest that Aristotle's defense of natural slavery stems from an inherited intellectual pressure to maintain a unified, hierarchical account of nature across the soul, household, and state.

INTRODUCTION

One of the most controversial topics in Aristotle's works is found in *Politics*, where he introduces the concept of slaves "by nature" (1255a6) and attempts to justify it philosophically. It has generated immense criticism, with scholars arguing against the ethical and moral shortcomings and implications of his assumptions. Explanations for why he espouses such views have been historically and culturally contextualized and debated by countless scholars. However, in this paper, I argue that even without using a moral or ethical argument against his theory, and working without knowledge of the socio-cultural and historical context of his time, Aristotle's account of slavery is incomplete—his argument hinges on the implicit assumption that behavior reliably reveals rational capacity in humans. Firstly, to understand Aristotle's perspective on slavery, we must read *De Anima*, for it offers us an explanation of why he initially suggests that the souls of slaves can be identified through their bodies. Secondly, we see that in the discussion of slavery in *Politics*, Aristotle admits that bodies are not reliable indicators of the type of soul one has, so he turns to observations about human behavior under tyrannical regimes to determine soul type. Yet his concepts of actuality and potentiality in *De Anima* and habituation in *Nicomachean Ethics* do not offer reasonable grounds for him to claim that slaves lack the deliberative part of their souls.

To preface, it is important to clarify that Aristotle's claim to natural slavery should not be understood as the assertion that slavery as an institution is natural. Aristotle distinguishes between slavery "by convention" versus slavery "by nature" (1255a6). What he means by *conventional slavery* is the enslavement

of war captives. He acknowledges that many are enslaved by agreement or law, not by nature, and that this arrangement is unjust (1255a1-1255b15). *Natural slavery* is a new category he invents in *Politics* to rationalize the enslavement of foreigners. Thus, the tension I am testing is whether Aristotle can coherently justify the concept of a *natural slave* at all.

SLAVERY IN POLITICS

Aristotle's main project in *Politics* is to identify the different forms of government and determine which is the best suited to enable humans in a community to achieve the highest good. His method is to break down the components of the state and study individual households, for he believes that the state is composed of many households (1253b1-2). Thus, his understanding of the relationship between different political institutions is largely reliant on his understanding of the relationship between men, women, slaves, and children within households. He claims that "there must be a union of those who cannot exist without each other; namely, of male and female, that the race may continue [...] and of natural ruler and subject, that both may be preserved" (1252a25-32). His view that this type of relationship is natural and necessary leads him to claim that "master and slave have the same interest" (1252a34-35).

Aristotle's breakdown does not stop at the household; it continues to individuals themselves. He claims that all living creatures consist of a soul and body, and that "one is by nature the ruler and the other subject"—the soul rules over the body with a "despotical rule." Specifically, "the intellect rules the appetites with a constitutional and royal rule." But this relationship cannot be flipped around, and neither can the intellect and appetites be equals. For creatures in "evil" and "unnatural" conditions, or who have "bad or corrupted natures," their bodies "will often appear to rule over the soul." Additionally, "the equality of the two or the rule of the inferior is always hurtful" (1254a34-1254b10). Since this proper, hierarchical relationship between body and soul exists at the smallest scale, this should reflect the relationship on a larger scale—within households, and then between households and the state. The excellent ruler deserves their position of power because they have the excellence of the rational, while those who are ruled have the excellence of the irrational (1260a4-7).

But the way freemen rule over various members of the household is different, because slaves, women, and children lack different capacities in their souls. According to Aristotle, slaves lack the deliberative faculty (1260a9-15). This, however, does not mean slaves are animals.

Although slaves and animals are both used to fulfill the needs of life, they still differ because slaves can "apprehend" but not actually have reason, while animals completely lack apprehension of reason and simply obey their passions (1254b21-25). What this all means is that slaves can understand reason when others speak it to them, but they cannot think or reason on their own.

This lack of the deliberative faculty is what Aristotle uses to justify why slaves need masters—they need someone who has reason so that they can survive, for they cannot govern themselves. He claims that just like how women

are inferior to men, slaves are also not on equal footing with freemen; they are of the "lower sort" and it is "better for them as for all inferior that they should be under the rule of master" (1254b13-20). Meanwhile, masters need tools to help them maintain their positions in life and run their households. This is why Aristotle defines slaves as living instruments and possessions for "maintaining life" (1253b27-34)—they are living tools meant to be used for tasks such as daily chores or laboring in the fields to help generate wealth for their masters. This paints a picture of a symbiotic, mutualistic relationship between both master and slave; they need one another to survive. When this relationship is natural (when each individual is in their rightful position within the relationship), the dynamic is beneficial for both; "they are friends and have a common interest." If the relationship is unnatural (e.g., an individual with the soul of a freeman becomes a slave), it will be hurtful to both (1255b9-15). However, though they are mutually dependent on one another, they are still not equals. The slave is like a body part of their master and belongs to their master entirely (1254a9-16).

But how do we identify slaves? How can we tell that someone lacks this deliberative faculty in their souls?

Aristotle initially turns to the physical body for indicators—this move will be further explained in my later section examining *De Anima*—and states that nature "would like to distinguish between the bodies of freemen and slaves." For example, slaves would be born with strong bodies so that they can be used for "servile labor," while freemen possess bodies that indicate they are useful for "political life in the arts both of war and peace." However, he immediately admits that this is not a reliable identification method: "But the opposite often happens—that some have the souls and others have the bodies of freemen" (1254b24-31).

This is why he subtly turns to another indicator to identify slaves: behavior. He uses the example of barbarians and the constitutions of the government they live under as his primary case study. Aristotle writes that "among barbarians," there is no real distinction between men and women in their communities, because "there is no natural ruler among them" (1252b5-9). He claims foreigners are "more servile in character than Hellenes" because they do not "rebel against a despotic government" (1285a19-21). But Aristotle also writes that these tyrannies rule over "involuntary," unwilling subjects (1285a26-28). So why do foreigners not resist? Aristotle claims that some foreigners, like the "natives in Asia," lack the spirit to resist and overthrow domination. Hence, they are always in a "state of subjection and slavery" (1327b23–33), with "no danger of their being overthrown," because "the people are by nature slaves." In fact, it is "hereditary and legal" (1285a21-23). In contrast, "No freeman willingly endures such government" (1295a23). Under the framework of how he understands individuals, households, and the states, along with the type of political and social organization that foreigners live under and seemingly endure without resistance, Aristotle uses this behavior as an indicator of what type of souls foreigners have. He believes they must have the souls of slaves because this is the only way he can make sense of their persistent subjection under despotic rule.

SCHOLARLY CONSENSUS

Aristotle's theory of natural slavery has generated an onslaught of criticism and disappointment among scholars. Many compare these parts of *Politics to Nicomachean Ethics* and use language like "incoherent," "very weak," "feeble," and even "battered shipwreck" to describe his theory of natural slavery to emphasize the "perceived logical incompatibility with the rest of his moral philosophy" and most "scholars' dismay at Aristotle's failure to recognize it as such" (Monoson 133). For instance, Nussbaum discusses how Aristotle's hierarchical views are incompatible with his ethical theory in *Nicomachean Ethics*. She argues that if all humans have rational capacities that can be developed through education, then his exclusion of women, slaves, and disabled people prevent their ability to develop their own capacities (Nussbaum 108-12).

Even after historical and cultural contextualization, such as recognizing how Aristotle's theory is a result of his "culturally specific experiences" (Monoson 134), scholars still see this theory as a "prime example of how even the finest mind can be corrupted by class interests, ethnic or race prejudice, or just the common orthodoxies of his time" (Monoson 133). For example, Schofield believes that Aristotle's justification reflects his bias as a member of a slave-owning culture and that he exhibits a "nasty case of false consciousness" (Monoson 136).

Kraut points out that Aristotle's theory is heavily supported by his belief in Greek superiority (Kraut 290-95), which is similar to other scholars who claim Aristotle is racist and ethnocentric.

In contrast to purely moral critique or historical contextualization, other scholars have focused on approaching Aristotle's theory by diagnosing its internal tensions, correcting anachronistic interpretations, or explaining the conceptual role the theory plays within his broader political project. Ambler discusses how Aristotle's theory of natural slavery is not consistent with actual slavery; it is impossible to ground political institutions purely in nature (Ambler 390-91). Instead, it exposes the tension between the failure of nature to match reality, and between nature and law (Ambler 399-402). Goodey argues that modern readers misinterpret Aristotle's theory because they project modern concepts of psychology such as cognitive disability (Goodey 203-4) and biological determinism into their reading of *Politics* (Goodey 210-12). Although Aristotle does seem to acknowledge the possibility of deformities and incomplete forms in *De Anima* (432b21–22), this understanding is not equivalent to the modern concept of intellectual disability, nor does it play an important role in his account of natural slavery. If such cases were relevant, we would expect them to be integrated into his political classification—but they are not.

Similarly, Dobbs criticizes modern, anachronistic interpretations of Aristotle, arguing that we must understand Aristotle's conception of slavery as teleological, not biological (e.g., Aristotle is not saying that natural slaves are mentally disabled by birth) (Dobbs 71-2). He attributes the failure of slaves to have full deliberative capacity to cultural practices and habituation through law (Dobbs 72-7). Kraut

believed that Aristotle had sufficient reason to defend the institution of slavery; to prove that the Greek *polis* (specifically Athens) was a "viable," "natural," and "ideal" political formation, he also had to explain why slavery was justified, as slavery was an important part of Athenian life (Monoson 137).

However, across the literature, I have yet to identify a discussion focused on how the inseparability of body and soul from *De Anima* may explain why he offers the initial suggestion that bodies can be identifiable markers of an individual with the soul of a slave. Furthermore, while several scholars like Nussbaum and Dobbs have emphasized the importance of the role of education or habituation in the development of rational capacities, their discussions do not engage with Aristotle's conceptions of actuality and potentiality that he articulates in *De Anima*, which I believe are especially relevant to the discussion of his theory on natural slavery. Thus, the rest of my paper aims to fill this gap by reading Aristotle's account of natural slavery through his theory of the soul, arguing that his distinction between first and second actuality, along with his account of habituation in *Nicomachean Ethics*, undermines his assumption that observing behavior is a sufficient method to evaluate deliberative capacity in one's soul.

ARISTOTLE'S THEORY OF THE SOUL IN DE ANIMA

As we have already seen in *Politics*, Aristotle's theory of natural slavery and how households and states are organized are both heavily reliant on his conception of the soul and its relation to the body. Although he does not discuss the soul in depth in *Politics*, he does so in *De Anima*. Specifically, in Book II of *De Anima*, Aristotle begins his project of defining and explaining the soul by discussing actuality and potentiality. He writes, "Now matter is potentiality, form actuality; and actuality is of two kinds, one as e.g. knowledge, the other as e.g. reflecting" (412b9-10). To understand his metaphor, I will use the example of wood. A block of wood, before it has been processed into something else, such as a wooden chair, a wooden spoon, or a wooden plank, faces many different possibilities in terms of its outcomes. Until its potential to be a chair, spoon, or plank has actually been realized, we can only discuss its potential uses.

However, after it has been formed into something or taken up another form—let's say a spoon—we can say that it has now reached the first kind of actuality and also a second kind of potentiality. Like acquired knowledge which remains latent until its active application, this spoon now has the potential to be used to scoop things (e.g., soup, rice, crafting beads). After the spoon realizes one of its potential outcomes—let's say scooping soup—we can say it has realized its second potentiality and reached its second kind of actuality.

Similarly, this is how Aristotle understands the soul. Firstly, he defines life as "self-nutrition and growth and decay" (412a13-15). Secondly, he makes it clear that the soul is not a body. Rather, the body is like matter or a block of wood, while the soul is like form or a wooden spoon; the soul is "the form of a natural body having life potentiality within it" (412a17-22). He continues on,

describing the soul as "an actuality like knowledge" (412a17-25). Thus, a soul is what enables an individual to carry out life-activities—it gives one possession of knowledge that allows one to perform self-nutrition (feed oneself), grow, and die. Thirdly, since the soul is what gives a body its form, it cannot be separated from the body. It is what organizes the body. Fourthly, the body is structured for the sake of the soul's activities (415b15-20). For example, the soul directs the body to move—it is the "original source of local movement" (415b22).

But not all souls are the same. Aristotle differentiates between the souls of plants, animals, and humans. Although the nutritive faculty or part of the soul exists in all living things (415a23-25), plants do not have what animal and human souls have—the perceptive faculty. This faculty enables the "possession of sensation" (e.g., touch) (413b1-7). Finally, humans are different from both plants and animals because they possess a third faculty—the rational or "calculation and thought" (415a7-10).

The implications of Aristotle's theory on the soul explain why he emphasizes the soul in *Politics*; if form and function are inseparable, and bodily organization reflects what the soul of the being is meant for, then human bodies, as organized by their souls, are "suited" for certain functions. This is why he initially suggests that we can look at bodies to determine the type of soul an individual has—the body is a reflection of the soul's capacities. However, as we have previously discussed, Aristotle already admits in *Politics* that bodies are not reliable indicators, which is why he turns to observable behavior. But the problem with this is that the logic behind Aristotle's conception of the soul—actuality and potentiality—proves that capacity is not equivalent to action; inaction does not indicate a lack of capacity or even a lack of ability.[135]

Possessing an ability does not require continuous, explicit manifestation of it to determine or prove that you have it. The same holds for capacity. Just because you have not expressed an ability to perform a certain act, it does not prove that the capacity does not exist. For example, I could have the knowledge of how to sing, but that does not mean that when I don't sing, I don't have that ability to sing. The knowledge and ability of singing already exists; I have already met the first actuality and now have the second potentiality for it. It merely lies latent. This also does not mean that just because I have yet to learn how to sing, I can never learn to sing. How can you know the capacity for that ability does not exist until you have had the chance to learn and attempt to develop that ability?

Similarly, I could have the capacity to deliberate or apprehend reason, or even already have that ability. But if I am never given that opportunity to develop that capacity or I never explicitly exercise that ability, does that mean it does not exist? When we apply this to slaves, we see that Aristotle's alternative solution

---

135 Note that here I use "capacity" to refer broadly to potential states—both first and second potentiality—and "ability" to refer to a developed state that can be exercised. An ability is simultaneously first actuality and second potentiality because the developed capacity now enables the subject to have the knowledge to act or realize the stage of second actuality. An easier way to understand "ability" is to understand it interchangeably with "knowledge."

for identifying the souls of slaves falls apart within the logical framework of his own theory of the soul. We can never really know if a slave lacks the capacity to deliberate if they are never offered the opportunity to express that capacity. Thus, Aristotle's mistake is that he collapses second potentiality into second actuality. He presumes that if someone possesses the ability or knowledge to perform an act, they will necessarily actualize or perform it. But why should this be a natural assumption?

HABITUATION IN *NICOMACHEAN ETHICS*

Habituation is the second necessary point of discussion, because it is the process required to move an individual from the first actuality of having the capacity for deliberation to the second actuality of active deliberation.

As Nussbaum and Dobbs have already suggested, habituation can shape human rational capacities. Nussbaum points out that exclusion and coercion prevent the development of capacities, while Dobbs argues that slaves are made, not born, because language acquisition, political education, and moral habituation can drastically shape one's nature. The reason they argue for this point is that Aristotle himself recognizes in *Nicomachean Ethics* that we become virtuous not by birth alone, but also through direct education and experience—"we are adapted by nature to receive them [intellectual and moral excellences], and are made perfect by habit" (1103a23-25). He goes on to state that "of all things that come to us by nature we first acquire the potentiality and later exhibit the activity" (1103a25-26), but that we learn things by doing them repeatedly. He uses the example of a builder; a builder learns how to build by building (1103a30-1103b1). What this means is that behavior is a reflection of training, rather than a reflection of mere capacity. Within the context of a political community, laws are meant to train good habits in us so that we can obtain the highest good for all within it.

As Aristotle has observed, foreigners live under tyranny. But the problem is that he treats the absence of overt resistance to tyranny as evidence for a lack of deliberative capacity—he conflates the *conditions for the exercise of political deliberation* with the *presence of political deliberative capacity*. This inference is unjustified. Additionally, he makes another mistake; he now collapses the distinction between capacity and ability, inferring the absence of capacity from the absence of its exercise (as we have examined in the previous section). His train of thought appears to be the following: A "natural" slave lacks the capacity to deliberate because they do not display the ability to deliberate. For if they truly possessed that ability—which would logically entail that they had that capacity in the first place—they would necessarily exhibit it in action.

Yet why should either of these assumptions hold?

In the first place, political deliberation does not necessarily result in outward acts of rebellion. There are many reasons for individuals to privately discuss and think about domestic political issues, yet not act on their judgments, given the great costs attached to political dissent under coercive regimes (e.g., fear of

losing one's life for rebelling against the regime, losing one's livelihood, or losing loved ones). The decision not to rebel may be a *reflection of prudent, rational deliberation, rather than evidence of its absence*. Furthermore, if foreigners are systematically oppressed and denied the opportunity to participate in the political sphere, they cannot meaningfully develop their deliberative capacity or publicly exercise existing deliberative ability, even if they possess it.

One might object that Aristotle is not relying on a single instance or case of inaction, but rather on what he has observed as consistent patterns of group behavior over time. However, persistent non-resistance under coercive conditions is still not a reliable indicator of a lack of deliberative capacity within a population, because such conditions suppress all incentive and opportunity for its expression.

Secondly, Aristotle also writes in *Nicomachean Ethics* that "We deliberate about things that are in our power and can be done" (1112a30-34). Under tyranny, the realm of actions within our power is extremely limited; if there are no political actions one can take without drastic consequences, what is there to deliberate about? Furthermore, Aristotle recognizes that slaves are not animals. This means that slaves are still considered human, who should have the rational part of the soul. But to say that the deliberative capacity within the rational part of their soul is missing, strictly through observations about tyrannical forms of government of these foreign slaves, is unjustified, for the development of deliberative capacity and the expression of deliberative ability could be entirely restricted by external circumstances.

One might argue that habituation under tyranny does not merely suppress the expression of deliberative capacity, but that the nature of the regime permanently renders the population incapable of it, since the people are never educated, trained to think politically, and practice deliberation. However, this would reflect the effects of political conditions rather than a natural deficiency. Aristotle's claim, by contrast, requires that natural slaves lack deliberative capacity *by nature*, not as a result of their circumstances.

CONCLUSION

Working solely within the logical framework in Aristotle's texts shows us that we do not need to rely exclusively on moral and ethical critiques or historical and cultural contextualization to understand where Aristotle's position on natural slavery stems from and how we can disprove his theory. His theory itself is incomplete and fails to hold up against the pressure of other theories (on the soul and on habituation) that its theoretical foundations rely on. His distinction between first and second actuality, together with his account of habituation and deliberation, undermines his assumption that the absence of deliberative action under coercive, tyrannical political conditions can serve as reliable evidence of a lack of deliberative capacity by nature—the conditions that enable deliberation are systematically destroyed by tyranny. The next step would be to figure out why he still chose to make this conclusion. But this is a

task that scholarship has already undertaken—countless scholars have already situated his view within the historical and cultural conditions that shaped his ethnocentric outlook. However, I would like to offer a different guess to why this is. Instead of claiming that Aristotle needed to justify the institution of slavery to support the idea that Athens had the ideal political constitution, as Kraut had theorized, I propose that Aristotle had to rationalize the institution of slavery because of his ambitious goal to have a coherent, consistent worldview that is true to what he deemed as natural through his observations. This is especially apparent in his desire to scale the "natural" relationship between individuals in a household to the state-level, and the way he needs his claim about the missing capacity of deliberation in the souls of slaves to justify his belief in the hierarchical relationship between the "ruled" and "ruler." He genuinely believes that the logic smaller units naturally operate on is the same logic that larger units should also be operating upon.

This is strikingly similar to what Plato also does in the *Republic*. Both teacher and student seek to use an account of our internal psychological order to explain how political communities should be naturally constituted. Thus, Aristotle's theory of natural slavery may not just be a product of his historical and cultural background, but the pressure of maintaining a unified account of nature across soul, individuals, households, and the state—a pressure inherited from the intellectual project and legacy his teacher had left him.

BIBLIOGRAPHY

Aristotle. *The Complete Works of Aristotle*. Edited by Jonathan Barnes. Revised Oxford Translation. 2 vols. Princeton, NJ: Princeton University Press, 1984.

Ambler, Wayne. "Aristotle on Nature and Politics: The Case of Slavery." *Political Theory* 15, no.3 (1987): 390–410.

Dobbs, Darrell. "Natural Right and the Problem of Aristotle's Defense of Slavery." The Journal of Politics 56, no. 1 (1994): 69–94. https://doi.org/10.2307/2132346.

Goodey, C. F. "Politics, Nature, and Necessity: Were Aristotle's Slaves Feeble Minded?" *Political Theory* 27, no. 2 (1999): 203–24.

Kraut, Richard. "Politics I: Slavery." In *Aristotle: Political Philosophy*. Oxford: Oxford University Press, 2002. Online edition, Oxford Academic, October 31, 2023. https://doi.org/10.1093/oso/9780198782001.003.0008

Lockwood, Thornton C. Review of *Nature and Justice: Studies in the Ethical and Political Philosophy of Plato and Aristotle. Aristote, Traductions et Etudes*, 38, by David Keyt. Bryn Mawr Classical Review, n.d. Accessed December 1, 2025. https://bmcr.brynmawr.edu/2017/2017.11.02/.

Miller, Fred. "Aristotle's Political Theory." In *The Stanford Encyclopedia of Philosophy*, Fall 2022, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University, 2022. https://plato.stanford.edu/archives/fall2022/entries/aristotle-politics/.

Monoson, S. Sara. "Navigating Race, Class, Polis and Empire: The Place of Empirical Analysis in Aristotle's Account of Natural Slavery." In *Reading Ancient Slavery*, edited by Richard Alston, Edith Hall, and Laura Proffitt, 131–150. London: Bristol Classical Press, 2011.

Nussbaum, Martha C. "Aristotle, Politics, and Human Capabilities: A Response to Antony, Arneson, Charlesworth, and Mulgan." *Ethics* 111, no. 1 (2000): 102–40. https://doi.org/10.1086/233421.

Plato. Plato: *Complete Works*. Edited by John M. Cooper. Indianapolis: Hackett Publishing Company, 1997.

"Signs of Physiognomy in Aristotle – Research Bulletin." Accessed December 1, 2025. https://research-bulletin.chs.harvard.edu/2010/11/30/signs-of-physiognomy-in-aristotle/

Vergouli, Vasia. Review of *Aristotle, 'Politics': A New Translation. The New Hackett Aristotle*, by C. D. C. Reeve. *Bryn Mawr Classical Review*, n.d. Accessed December 1, 2025. https://bmcr.brynmawr.edu/2018/2018.01.12/.

Abigail So is a senior at Cornell University from Hong Kong, double-majoring in Classics and Government, with a minor in History. Her studies primarily focus on Western political theory and intellectual history, but she is also interested in Chinese foreign affairs, Egyptology, classic literature, religious and cultural syncretism, and applying psychoanalysis to the study of politics. When she is not arguing with dead philosophers or hyperfixating on a strange internet niche, she enjoys producing beats, singing, writing lyrics, dancing, playing video games, and learning new languages.

AUTHOR'S NOTE

Aristotle's remarks on slavery in Politics first came to my attention during my first semester at Cornell, while I was sitting in my Ancient Greek history class with Professor Barry Strauss. I should've reared back in disgust or disappointment, but instead, I found myself morbidly fascinated. How could such a highly regarded ancient philosopher—one known as the "father" of logic—come to such a conclusion? The question slipped to the back of my mind until it re-emerged in my junior year, during a reading of Politics with Dr. Ani Chen for her Ancient Political Thought course. This time, I noticed that Aristotle heavily emphasized the "souls" of individuals. Unfortunately, I was ill-equipped to understand what he meant by "soul"—my understanding derived from my Christian upbringing—and surmised that he was merely biased against foreign "barbarians."

Yet a part of me still felt this explanation was inadequate. To put my sense of aporia to rest, I chose to spend the fall of my senior year studying Greco-Roman philosophy with Professor Ted Brennan and reading through a selection of Aristotle's works with Professor Rachana Kamtekar. The moment I finished reading De Anima and Nicomachean Ethics for both classes, it was clear to me that Aristotle's tendency for logical consistency across his works fell short

in the face of his natural slavery argument. Determined to work through this issue so that I could explain to myself—and to others—why this was so, I took the opportunity to write about this for my final paper in Professor Kamtekar's class. I would like to thank her for her kind suggestion to submit this work to an undergraduate journal. Without her encouragement, I would have never considered my ideas worthy of publication.

Making Sense of Mental Ballistics

Uriel Vargas
University of Chicago

FOREWORD

Editors: Corbin Breeden, William Liaw, Maria Lim, Will Taylor, Sammi Wu

Few philosophical questions are as immediate—or as unsettling—as whether we truly think, or whether thoughts simply happen to us. When we reason our way toward a conclusion, are we active participants or mere observers of the operation of our minds? "Making Sense of Mental Ballistics" confronts this question through an engaging and critical examination of Galen Strawson's skepticism about mental action, asking whether reasoning can genuinely count as intentional action.

What makes this paper especially compelling is its rejection of a false choice between total control and complete passivity. Drawing on the work of Wayne Wu, Ludwig Wittgenstein, and contemporary action theory, Uriel Vargas argues that reasoning can remain intentional even when some of its constituent processes unfold automatically. This result is a nuanced and persuasive account of agency — one that presents a substantial challenge to Strawson's rejection of mental action as an intentional process.

We were particularly impressed by the paper's clarity, rigor, and originality. It tackles a difficult and highly technical debate while remaining accessible and deeply relevant. After all, our practices of inquiry, responsibility, and self-understanding depend upon the idea that we are, at least sometimes, authors of our own thoughts. By showing how intentional agency survives amidst mental spontaneity, Vargas offers a rich and illuminating contribution to contemporary philosophy of mind and action. It invites readers to think more carefully about what it means not merely to have thoughts, but to *think*.

ABSTRACT

If intentional actions are events that are consciously initiated and controlled by agents as they unfold in time, Strawson argues that little of what we consider to be "mental action" is intentional action. Reasoning, so Strawson argues, is not immune from this skepticism. While agents can initiate the process of reasoning, they do not control its unfolding over time. In this paper, I challenge Strawson's skepticism and argue that reasoning toward a judgment can be an intentional action. I argue that when we reason toward judgments, an agent's sustained attention, coupled with their interaction with thought content, are exhibitions of agential control.

Furthermore, I argue that a rigid understanding of control as completely continuous control greatly limits our capacity to act intentionally with our bodies and our minds. For this reason, intentional action's control criterion ought to make room for instances of automaticity.

INTRODUCTION

A near ubiquitous philosophical assumption is that reasoning is something agents do. So ubiquitous, in fact, that we embed it within our conceptions of agency and moral responsibility. However, Strawson asserts that when we reason toward a judgement, our reasoning cannot, in any sense, be understood as a matter of action. This surprising conclusion, which consequently undermines many of our agentive and ethical concepts, depends on a distinct characterization of intentional action, one that construes it as a special type of event that necessarily exhibits two features: 1) the event is consciously initiated by an agent and 2) between the event's initiation and termination, the agent controls the event's temporal unfolding. According to Strawson, an agent can intentionally initiate the process of reasoning toward a judgment, but the process does not meaningfully unfold under the agent's control such that they intentionally reason up until the formation of a judgment.[136] Not only is there strong phenomenological support for this conclusion, but, so Strawson argues, the rule-bounded structure of reason itself constrains inferential transitions in a way that after initiation one simply hopes that rationality operates "properly and productively."[137]

In this paper, I argue that Strawson's mental action skepticism is mistaken. I show that agents can intentionally reason up until the formation of a judgment by meeting the control criteria in several ways. I incorporate and modify Wu's remarks on attention to show that an agent's continuous and sustained attending is a manifestation of the agent's intention to form a truthful belief bound by the rules of reasoning. Furthermore, agents interact with thought content in a manner consistent with their continued intention to terminate a process of reasoning to a judgment. Both attention and interaction with thought content are, properly understood, exhibitions of agential control throughout the unfolding of such a process. I end the paper by showing that a rigid understanding of the control criterion as completely continuous control, as "through and through control," greatly limits our capacity to act intentionally, not just with our minds, but with our bodies. As such, intentional action's control criterion ought to make room for necessary instances of automaticity.

In section one, I motivate Strawson's broad mental action skepticism and explain that the case of reasoning toward a judgement is not immune from it. In section two, I formalize Strawson's argument by constructing a model of reasoning toward a judgement, strictly adhering to his terminology and ideas. In section three, I apply Strawson's model and make concessions about thought content coming to mind immediately after initiating the process. Thereafter, I begin to show that agents exhibit control by incorporating Wu's insight that agents sustain their intention to Φ through their continuous attention. In section four, I refute the argument that the rule-boundedness of reason necessarily

---

136 Strawson, (2003), 231-234, 247-248
137 Ibid., p. 233

implies agential presence, but not agential control and show that agents do what I call "Taking Stock" or "Unification". In section five, I show that a control criterion in intentional action ought to make room for instances of automaticity and that intentionality and control can exist on a continuum.

I. SETTING THE STAGE: MENTAL ACTION SKEPTICISM

Strawson's skeptical challenge rests on a persistent feature of our mental lives, particularly, that thoughts impinge on us.[138] One might be sitting in a lecture hall, uninterested in the material being presented, before realizing that the mind is off to the races, now producing thoughts about dragons, what to cook this week with the groceries in the fridge, and how to prove that pesky logic problem without running into contradictions. Experiences of daydreaming and intrusive thinking, the two most paradigmatic cases of thought-impingement, are both characterized by two features: the absence of agential control as either process unfolds and that we often find ourselves amid them without any internal, conscious initiation. During both events, the agent has the experience of thoughts happening to them, uncontrolledly producing themselves in succession without need for the agent's conscious initiation to get either process going. As such, in the absence of a guiding intention, an initial belief/desire pair to nondeviantly cause the action[139], or any degree of control over the manifestation of the event, I take it that most philosophers of action would agree that both paradigmatic cases of thought-impingement are merely mental happenings rather than mental acts.

Since an event that exhibits both features that characterize paradigmatic cases of thought-impingement is not action, one might then use the features as negative criteria to characterize mental acts. That is, an event is not merely a mental happening, but a mental act if and only if a mental event is both 1) intentionally initiated by an agent and 2) unfolds under the agent's control. We have good reason to make this assumption. In fact, all we would be doing is applying the criteria of our more robust theories of intentionally acting with our bodies to the domain of mental action.[140] However, if we do so, Strawson poses a skeptical challenge: little, if any, of what we consider to be "mental action" fulfills the control criterion. Accordingly, he argues that little of what we consider to be mental action is intentional action.

Consider the following: 1) *remember* the last ice cream flavor you had 2) *imagine* a starry landscape, and 3) *divide* 100 by 5. If you read the sentences, understood what the imperatives asked of you, and gave them your full consideration, Strawson argues that the difference between these set of mental

---

138 Ibid., p. 229

139 Davidson (2001, p. 79). For more on deviant causal chains, see Piñeros Glasscock & Tenenbaum (2003, Section 2.4: The Problem of Causal Deviance).

140 I have in mind Davidson's and Frankfurt's theories of intentional action. Requirement one explains that the cause of the action originates from an agent's internal dispositional state, or, as Davidson would put it, a primary reason. Requirement two inherits Frankfurt's insight that intentional actions are better understood by their sustaining causes, that is, the agent's continual guidance and control. See Davidson (1963) & Frankfurt (1978).

events and the paradigmatically thought-impinging kind is only an intentional act of initiation or catalysis.[141] In both sets of events, the agent fails to meaningfully exercise any control over the way the event unfolds.

Strawson's conclusion is surprising, and if it seems implausible, consider the explicitly non-agentive descriptions that might appropriately capture the experience of remembering, imagining, and completing computations. It might have "struck you" that the last pint of ice cream you had was Honey Lavender, "come to you" that 100 divided by 5 was 20, or imagined the starry landscape by simply waiting for your brain to project the image onto your mind.[142] Strawson's picture of mental agency posits that after intentional catalysis, an inactive agent simply waits for their neurons to fire in the right way to appropriately create and spring up thought-content to address the issue raised by their catalysis.[143] All cases are marked by remarkable automaticity. Importantly, unlike the remarkable physical control that agents seem to possess, through and through, when we slide cups across a table or raise an arm, mentally, all that the agent seems to be doing is intentionally triggering an uncontrolled onslaught of thoughts. Thus, since the agent fails to meet both the initiation and control criteria, the agent does not, according to Strawson, remember, compute, or imagine intentionally. These activities are, at best, mere effects of something we do intentionally, but not themselves things done intentionally since their unfolding is not under our immediate control.

We can begin to appreciate the consequences of Strawson's skeptical challenge once we consider mental events other than recall, imagination, and computation. Consider the case of reasoning toward a judgement, which is not immune to the skeptical challenge.[144] When one attempts to form a belief concerning a difficult problem, for example, there is, firstly, the initial setting of one's mind to make a judgement about some problem. I, for one, have the habit of physically signaling my internal act of catalysis by looking up and to the left, or by lightly resting the thenar webbing of my left hand over my mouth. But after one's intentional act of initiation, there is a sense in which one simply waits for the brain to execute its task, in the same way that one does when one remembers, imagines, and computes.[145] When the time comes to produce a judgement for the mental matter at hand, it seems that all we do is hope our brains produce the appropriate thought-content (a theorem, a quote, or a statistic) and the right connections between the relevant thought-content (the relations between propositions, the quote's relevance to a key idea, or how the statistic was computed).

Skepticism about the extent of our agential control when we reason toward a judgment goes beyond the strong phenomenological passivity that resembles the experience of imagining, computing, or remembering. Strawson argues that

---

141 Strawson (2003, p. 231)
142 Ibid., pp. 229-230
143 Ibid., pp. 231-238
144 Ibid., pp. 231-239, 247-248
145 Ibid., pp. 233-234

if we are agents that are also reasoners that *strictly* adhere to the rules that structure reason itself, then rule-bounded agents do not have complete control of the inferential transitions they can make.[146] The inferential transitions available to the agent are only those that are licensed by the rules of reasoning. If all X's are Y's, and all Y's are Z's, the agent cannot, for example, choose to deny that all X's are Z's. Or, if a reasoner truly knows that P implies Q and they see that P, the reasoner is forced to conclude Q. So, if all processes of reasoning toward a judgement are composed of many of such inferential transitions where rule-bounded agents cannot control the transitions between thought content, and if they are strictly adhering to the rules that structure the operation of reason, then the agent fails to meaningfully control the unfolding of the mental event. The agent can intentionally initiate the process of reasoning to form a belief, but if they are bound by the rules of reasoning, then the agent, merely waiting for the "natural causality of reason to unfold" and its "proper and productive operation," fails to intentionally reason toward a judgment. [147]

Strawson's conclusion that reasoning toward a judgment is not a matter of action is worrying. Many of our agentive and ethical concepts rest on the assumption that reasoning toward judgements is something we do. A common conception of agency, following the influential work of Anscombe and Davidson, is that agency involves the capacity to perform intentional actions based on the capacity to act for a reason, where acting for a reason involves "[acting] in a way that can be rationalized by the premises of a sound practical syllogism."[148] Rationalization, the exercise of reason to deliberate about our goals, their desirability, and our means to attain said goals, which results in an action, is, under this view, the site of our agency. So, if reasoning were not an action, if the entire process of rationalization merely impinged on us and occurred with the same automaticity as, say, digestion or respiration, agency disappears entirely.[149] We would merely be what Strawson calls "Pure Observers": well-functioning and "cognitively well-equipped" automata that are mere hosts to "trains of reasoning" that lead certain beliefs and hypotheses to impinge on us; "reasoning, thinking, judging self-conscious creatures" that lack agency and are incapable of intentional action.[150]

The assumption that we could be Pure Observers complicates our commonsense conceptions pertaining to our moral responsibility. Beings are held morally responsible not merely on the grounds that they cause certain actions, but that they cause certain actions even though they possess certain properties, capacities, or powers. According to the reasons-responsiveness view of moral responsibility, the relevant powers lie within the capacity to exercise consideration for rational criteria; in seeing reasoning and rational deliberation

---

146 Strawson (2003, p. 233)
147 Ibid., pp. 232-233,
148 Schlosser (2019).
149 Alvarez (2009) makes a convincing argument that it is possible to perform an intentional action without having to generate reasons by reasoning. However, there is an admission that these cases "tend to have little significance" and that, perhaps, paradigmatically intentional acts are those done for reasons.
150 Strawson (2003, pp. 228-229)

as something agents do that tangibly bears on their actions.[151] The assumption that reasoning is not something agents do, in turn, threatens this common way of attributing moral responsibility for our actions. Because Strawson's skepticism threatens our agentive and ethical concepts, we need a resolution. The remarkable passivity that we experience after intentional acts of catalysis compromises the status of much of what we take to do with our minds.

II. BUILDING A MODEL

Strawson claims that intentional acts of catalysis beget helplessly passive agents who look up in the air in hopes that their minds, not them, will make sense of mental ballistics by springing up thought content in the right way.[152] Thus, to understand his claim, it is important to formalize it for clarity.

Suppose that a process of reasoning toward a judgement is denoted by Φ. Strawson articulates that there are three central components of Φ. Firstly, there is what gets Φ going, the initial moment of setting oneself to engage with any Φ. This might take the form of internally telling yourself, "Ok, no more messing around. Time to judge or reason about X" or, what might happen more often, that distinct, silent mental operation of orienting and devoting the mind's resources to the task at hand.[153] Strictly adhering to Strawson's vocabulary, let this moment be denoted by R_1 and call it the agent's intentional act of catalysis. An R_1 in Φ, Strawson argues, is roughly analogous to the moment when a sharpshooter intentionally fires a projectile. The sharpshooter is in control of initiating the process which results in the motion of the projectile, but after their intentional firing, the "rest is up to nature."[154] The sharpshooter's control ends at their intentional firing; they exercise no control over the projectile as it travels to the target.

In addition to an agent's R_1, a process Φ is composed of two additional components. There is the moment that you come to judge that X or judge that Y. Call this "the termination of Φ" and let it be denoted by Ω. Lastly, call the gap between R_1 and Ω "the evolution of thought content." In Strawson's words, the evolution of thought content is the "inferential [moving] between particular contents," "thought considered in its contentual essence— the actual confrontations and engagements between contents, the collaborations, competitions, [and transitions] between them" that occur after we R_1.[155] This gap is the moment when the agent, according to Strawson, does nothing but hope that the rational powers of their mind produces the appropriate contents and connections between them to meaningfully terminate Φ with Ω.

If we accept that these formal features accurately describe the central elements of Strawson's view of mental agency, then, within these parameters, he

---

151 McKenna (2022)
152 Strawson (2003, p. 234)
153 Ibid., p. 231
154 Davidson (1971). For more on ballistic action see Stout (2018).
155 Strawson (2003, pp. 233-234)

claims the following: R_1 is an intentional action, but in the progression up until Ω—the gap between R_1 and Ω—there is remarkable passivity where the rational powers of the mind operate properly and productively without the agent's control. R_1 is intentional, but the ensuing gap, the onslaught that successively introduces more of itself without agents themselves producing or controlling any of it, is, as he puts it, merely ballistic and uncontrolled. Therefore, since all we do is intentionally trigger the evolution of thought content by R_1'ing, when Ω terminates **Φ**, all the agent has done is intentionally R_1'ed but not intentionally **Φ**'ed.[156]

To regain agential control between R_1 and Ω, one might try to disambiguate the analogy between firing a projectile and engaging in a process **Φ** by considering the following scenario: Suppose that I ask you to judge whether some logical proposition is true. After you R_1 and diligently set yourself to make the judgement about the proposition, Emma makes a loud noise that interrupts your train of thought before your mind produces Ω. Since the process is yet to be completed, to get back on track to terminating **Φ**, one must, so to speak, "pick up where they left off" and set one's mind at the mental matter again, call it R_2. The case of firing a projectile has no such notion of a resumption, R_2, amidst the process of the projectile traveling to the target. After a freekick, for example, a soccer player cannot, during the process of the ball traveling to goal, "pick up where they left off." They must wait for their freekick to be unsuccessful, start over, and then once again R_1. This other initiation will not be a resumption of their original R_1, but a distinct R_1 that marks an entirely new process to be completed. Thus, one might say, since the case of reasoning toward a judgment exhibits features of agential resumption, unlike the case of hitting a target, the agent can control the action after R_1 in such a way that a reasoner can intentionally **Φ**.

However, we can extract a rebuttal from Strawson: even if there is an additional act of catalysis between your R_1 and Ω, you don't meaningfully participate any more in your action such that you intentionally **Φ**. All you have done is add an additional gap between R_1 and Ω (R_1 to R_2, and R_2 to Ω) but during both gaps, all you do is wait for your brain to produce the right content and the right connections between them to enable you to terminate **Φ**. Moreover, with the additional act of catalysis, the agent does not become free to control the inferential transitions between thought contents. We can introduce this as the final feature of Strawson's model: Suppose that between R_1 and Ω there <u>can</u> be an indefinite amount of catalytic primings R_n to get to Ω, where n is assigned a number by adding 1 from its preceding catalytic priming, but not in a way that the agent intentionally reasons to a judgment. Therefore, we can have '**Φ**: {R_1...R_2...Ω}' and '**Φ**: {R_1...R_2...R_3...R_4...Ω},' etc., without **Φ** ever being something we do intentionally. No matter how "regularly [we repeat]" our catalytic primings, no

156 I will not concern myself with the question of whether one can intentionally Ω. Our concern is the space between R1 and Ω. Unsurprisingly, Strawson also believes that Ω is not action. For more on this subject, see Hieronymi (2011), Setiya (2008), Vermaire (2021), Kelley (2024).

matter how much Φ is composed of "conscious primings," our Φ's are doomed to be dominated by large-scale automaticity.[157]

III. ATTENTION

Now, equipped with our more precise language, let's consider proposition P. In the process of determining whether P is true, we must also exercise our metacognition to determine whether Strawson is right to say that the agent helplessly waits for their mind, not them, to produce an Ω:

P: If β is contingent, then, for any ψ, either the argument from β to ψ is invalid or the argument from β to ¬ ψ is invalid.

There is a sense in which after the agent R_1's and understands precisely what they are setting their mind to, relevant thought content *comes* to mind, just like that, without the agent themselves picking it out. Regarding P, for example, the mind does not produce many thoughts about puppies, concertos, or ice cream immediately after an R_1. The mind produces, in higher abundance, thoughts specific to addressing proposition P for the sake of getting to Ω without the agent themselves picking it out.[158] So, the mind produces thoughts about what β and ψ might denote, or perhaps what the formal definitions of validity and contingency are. This process—the production of thoughts immediately after R_1—is automatic, and I conjecture, is the result of being the sorts of agents that must make judgments for their survival. Additionally, Wu astutely remarks that general automaticity in mental action stems from having limited brains that cannot, like a supercomputer in science fiction, consider massive amounts of inputs.[159]

Suppose that every time one tried to terminate any Φ, all the agent's thought content came to them at once. If so, the emergent thought content would be completely indiscernible and impossible to conceptualize. Additionally, terminating any Φ would also be impossible. The agent would be stuck weeding through muddles of thought content to eventually get to what is relevant for terminating the Φ at hand, hoping that they don't accidentally give their attention to some other piece of thought content they weed through and hopelessly restart the process all over again. We need the automaticity immediately after R_1 to put us in a position to terminate Φ, and the brain, limited but efficient, mostly

---

157 Strawson (2003, pp. 231, 234). Careful readers of Strawson may say I misunderstand the notion of an Rn by specifically construing it as an intentional act involving the devotion of one's mental resources by attending to process Φ. After all, he believes that there are many other catalytic acts like "silently imaging key words, rehearsing inferential transitions," dragooning a wandering mind, and receptive blankings that give "missing elements a chance to arise" (231-232). If there exists a moment that initiates Φ, then these other catalytic acts cannot possibly be understood as R_1's. As a simple matter of fact, one cannot image key words, rehearse transitions, dragoon, and initiate receptive blankings without already having attended to some content in mind. Are these other catalytic phenomena intelligible as Rn's after R_1? Strictly as secondary manifestations of the more basic act of attention, as ways for agents to interact with thought content consistent with their continuous and sustained intention to form a truthful judgment. See sections III & IV.

158 I am careful to leave room for imperfections. It is conceivable that after a Strawsonian R1 unrelated to puppies or ice cream, our minds still produce thoughts about ice cream or puppies. Whether immediately after R1 or somewhere further along as the agent gets closer to Ω. This is, after all, the nature of intrusive thoughts.

159 For a richer and far more scientific treatment of our cognitive capacities, see Wu (2013, pp. 246- 250).

produces thoughts about what is relevant to our R_1 to eventually reach Ω.

Despite this concession, our consideration of P should raise concerns that the commitments that constitute Strawson's model accurately capture the experience of reasoning to a judgement and that together, they explain that Φ is not an intentional action. We ought to revisit our earlier remarks about agential resumption, which imprecisely hint at a resolution. If an Rn is to be understood as an intentional act involving the devotion of one's mental resources, as an intentional act of attending, Strawson misapplies the sharpshooter analogy. When a basketball player releases a free throw, or when a soccer player kicks a penalty, it is true that they cannot exercise control over the projectile during its motion. They can adjust the movements of their bodies prior to their release or kick to ensure that the projectile lands where they intend, but the kick and the release of the ball are discrete temporal events that cause the unfolding of an event where, strictly speaking, the players are completely absent. Even though Strawson wants us to believe that they are similar, Rn's are unlike the kick and release of a ball. Why? Simple: attention is continuous, not discrete.

When one attends to a mental matter, one concentrates and holds the relevant thought content in mind, not just discretely at the moment when one wants to begin thinking about a certain mental matter or after one gets distracted, but continuously thereafter both sets of events. I take myself to be describing the nature of attention, the explicit and intentional devotion of one's mental resources, unproblematically—as all agents experience it. In fact, to crystallize the point that attention is, by necessity, continuous, it's also helpful to illustrate the strong relation of dependence between the successful execution and termination of several mental events and an agent's continuous attending, in all temporalities. Mental events are said to have occurred, be in the process of occurring, or set to occur sometime in the future, only if the agent has attended, is attending, or will attend to the relevant thought content necessary to execute and terminate the relevant mental matter. If an agent forgets to devote their mental resources to imagine, remember, compute, or reason toward a judgment sometime in the future, or if Emma's distraction is so monumental that they unrelatedly begin thinking about how delicious Grinch Salt is, then, logically and as a matter of fact, the agent has not Φ'ed, is not currently Φ'ing, or will not Φ.

Thus, an intentional event Rn is unlike the causal initiations involved moments before the motion of a projectile where the agent can merely effectuate the unfolding of an event where they are, strictly speaking, completely absent. Taking a penalty, a free throw, or firing a gun, whether successful or unsuccessful, all terminate after the unfolding of an event where the agent lacks the capacity to control the projectile amidst its motion, when the ball or bullet come to a stop. Contrarily, the execution of Φ depends on the agent continuously holding and concentrating on relevant thought content until the agent arrives at an Ω. Since the agent has Φ'ed, is Φ'ing, or will Φ in virtue of their continuous concentration, their attention controls the unfolding of Φ.

Wu's work on the primacy of attention enables us to substantially strengthen the position against Strawson's skepticism by explicitly incorporating

the notion of an intention within our discourse. Making our previous remarks on the nature of attention more robust, Wu argues that an agent's attention is a manifestation of their sustained and continued intention to terminate a mental process.[160] I deviate from Wu in explicitly construing attention as the emergent phenomenon that enables us to solve the Many-Many Problem.[161] By no means is his explanation implausible or incorrect, but the purposes of this paper do not demand a technical understanding of attention other than the phenomenological experience common to all agents. We have already argued that the act of saying that the agent has Φ'ed, is Φ'ing, or will Φ, is intelligible if and only if it is true that the agent is attending, has attended, or will attend to the mental matter. Thus, it's no stretch to incorporate Wu's insight and say that the agent's continuous and sustained attention is just a manifestation of their intention to form a truthful judgment in accordance with the laws of reason. Thus, our understanding of Rn's ought to admit of more nuance than what is available to us in Strawson's model. Rn's function as the beginning of the agent's continuous and sustained attending, but simultaneously, as manifestations of their continuous and sustained intention to arrive at an Ω that adheres to the rules of reasoning.

IV. RULE BOUNDEDNESS

However, our remarks on attention and our incorporation of Wu's insight do not completely diffuse Strawson's skeptical challenge. In fact, he anticipates our move to locate agential control in the unfolding of Φ through the agent's continuous and sustained attention:

> "Attention, too, can be a matter of action, of maintaining attention; but 'Attention creates no idea,' as William James remarked. In itself it delivers no new content, and it need not be a matter of action at all, any more than being keyed up and tensely expectant are. One may be gripped, fascinated, absorbed, swept away, one's attention may be held: all these descriptions correctly imply lack of action." (Strawson, 2013, p. 232).

The most robust form of Strawson's skepticism grants that the agent maintains their attention on thought content relevant to terminating Φ throughout the entirety of the process. However, it denies that the agent's continuous and sustained attention is an indication of their control throughout the unfolding of Φ. Strawson argues, in line with the idea of a Pure Observer that is merely a host to trains of reasoning, that sustained attention is just an indication of *agential presence*, but mere presence does not necessarily imply that the agent is *doing* anything. Attention simply allows us to spectate the operation of reason within ourselves.

There is no "intentional agency to the operation of reason" because reasoning is, by definition, a rule-bounded enterprise.[162] So long as agents

---

160 Wu (2013, pp. 253-256) & Wu (2022, pp. 62-65).
161 Wu (2022, pp. 65- 68) & Wu (2013, pp. 250-253).
162 Strawson (2003, p. 233)

adhere to the rules that constitute reasoning, they do not possess total control of the inferential transitions between thought contents. Thus, while it is true that agents sustain their attention all throughout Φ, the heavy lifting, so to speak, is done by the rational powers of the mind which operate in accordance with the rules that constitute its capacity. The agent, through their continuous attendance, "keyed up and tensely expectant," is but a passive spectator to the proper and productive operation of reason.

To challenge the most robust form of Strawson's skepticism, let's turn our attention to the game of chess. Chess shares many similarities with reasoning. Both activities are teleological, they unfold over an interval of time, and require intentional initiation and continuous and sustained attention all throughout the duration of both events. However, the relevant similarity that concerns us is that both are rule-bound enterprises. Just like reasoning prevents the agent from making certain inferential transitions, the rules that constitute the game of chess only license a particular set of moves. A player cannot move a bishop like a rook, a queen like a knight, or a pawn like a king. Moreover, given a certain position, chess also necessitates certain consequences. In reasoning, if all X's are Y's, and all Y's are Z's, the reasoner cannot choose to deny that all X's are Z's. Similarly, if a player's king is in check, they cannot choose to continue the game by executing a move that does not protect their king. As such, for both activities to reach their proper termination, we accept that the rules that constitute them are fixed and unmodifiable by the agent's volition. Else, if a reasoner or a player is not adhering to the rules that constitute the rule-bound activity, then they are not actually playing chess or reasoning. Given that chess and reasoning are rule-bound in the same way, is it also true that the chess player, or the participant in any other rule bound enterprise, does not control the unfolding of a chess game just as the reasoner does not control the unfolding of Φ? Absolutely not.

Even though a chess player is bound by the rules that constitute chess playing, the rules of chess do not happen to the player in such a way that they do not control the series of moves they execute. Contrarily, the player intentionally initiates the activity, sustains their attention all throughout, and chooses to pursue a particular strategy in accordance with their intention to win the game. Broadly constituted, chess is about considering sequences of moves, testing connections between these sequences of moves, and thereafter, choosing which sequence of moves to pursue in accordance with one's sustained intention to win the game. Just as a player can think of a bad move completely consistent with the rules of chess, a player can also think of a move inconsistent with the rules of chess. Thus, much of the skill involved in playing chess is exercising one's mental dexterity to determine **A)** which moves are consistent and inconsistent with one's intention to win the game as the game continues to evolve and **B)** which moves are licensed by the rules that constitute the activity. As such, the player can willfully discard the thought of a bad move from their consideration, focus only on those thoughts consistent with their intention to win, and continue the process of refinement throughout the game to inform the

execution of their moves. All of this is done in accordance with the rules of the game, but in no way does the rule-boundedness of chess prevent the player from controlling the game of chess as it unfolds.

In the same way that rule-boundedness does not prevent chess players from controlling the unfolding of a game of chess, the rule boundedness of reason does not prevent agents from meaningfully controlling the unfolding of Φ between R_1 and Ω. In addition to intentionally initiating Φ and sustaining their attention throughout the process, agents interact with thought content in accordance with their intention to form a truthful judgment informed by the rules of reasoning. In the most abstract sense, the interaction with thought content that moves the process of reasoning forward manifests in two different ways, that I call Taking Stock and Unification.

Taking Stock: Agents engaged in Φ go through their pool of thoughts and examine which ones are worth keeping, which thoughts are and are not consistent with their intention to arrive at a determinate Ω.

Unification: Only occurs *after* an agent has "Taken Stock" and has pool of thought-content consistent with their intention to get to an Ω. This is a process of apperception and interaction within the set of thoughts determined to be consistent with the agent's intention. Agents hold in mind two, or more, discrete elements within the set and test to see if the elements are related and can, in any way, be unified into new and novel abstractions previously unconsidered.

Just as a chess player exercises their mental dexterity to determine which moves adhere to the rules that constitute the activity and are consistent with their intention to win the game, agents engaged in Φ go through their pool of thoughts and examine which ones are consistent with their intention to form a judgment Ω. Thought content that is licensed by the rules of reasoning and adheres with one's intention to make a truthful judgment is kept within the agent's scope of consideration. Everything else is discarded and set aside. This is precisely what is involved in "Taking Stock." If we are trying to determine whether P is true, the thought that contingency means that a sentence in sentential logic is always false is unhelpful. So are the thoughts that an argument can be valid if it is sometimes possible for the premises to be true or that Grinch Salt is delicious. All these thoughts are inconsistent with our persistent intention to form a truthful judgment about Φ. So, a large part of terminating any Φ is exercising our mental dexterity to check which thought contents are worth keeping and which are not, and thereafter, placing them into another pool of thoughts which continue to get further refined until it is possible to arrive at a determinate Ω.

Furthermore, just as a chess player compares and tests connections between sets of moves to examine new and novel moves that might be consistent with their intention to win the game, a reasoner can compare thought contents and test connections within those thought contents to see if new and novel connections arise that might help them terminate Φ. This is what is involved in "Unification." After Taking Stock, an agent determines that two, or more, distinct pieces of

thought content adhere to the rules of reasoning and are consistent with their intention to terminate Φ, they can examine whether such contents they have in mind can be unified into entirely distinct pieces of thought content that might, in turn, help them terminate Φ. We might perfectly understand the components that constitute the consequent of P ("*for any ψ, either the argument from β to ψ* is invalid or the argument from *β to ¬ ψ is invalid*") but we won't get anywhere if we don't test the relations implied within it and the larger relation implied between the antecedent and the consequent. The thought that the relationship implied by P holds when *ψ* is a contradiction (or when it is contingent or a tautology), for example, is not something that arises spontaneously. Rather, it arises from the agent's examination and consideration of the thought content in mind, coupled with deliberately comparing distinct contents to unify them in new ways consistent with the rules of reason and our intention to make a truthful judgment.

To provide a rough physical analogy to what I have in mind when agents intend to terminate Φ, consider the following thought-experiment. Suppose that your mind and all its contents were represented in a box. Imagine that each piece of content in the mind is represented by an index card, and that index cards are sorted by color depending on the type of content that is on them. Content about ice cream is represented with red index cards, whereas content about logic will be represented by blue index cards. Rather than introspecting, terminating Φ depends on opening the box and physically accessing its contents, picking out index cards until one arrives at a determinate Ω. Within the constraints of the thought-experiment, Strawson's view of mental agency claims the following: the agent intentionally sets themselves to terminate Φ by opening the box; the agent intentionally sustains their attention on the index cards until they reach Ω; the agent merely observes their hands, informed by the rules of reasoning, pick out the relevant index cards from the blue pile to help them get to Ω.

Instead, I argue that the process is more agentially involved. Agents sustain their intention to terminate Φ by opening the box and interacting with the index cards inside until they reach Ω. Furthermore, because agents have limited but efficient brains and are the sorts of beings that must make judgments for their survival, initially, their hands know to go to the blue pile and pick out some cards that might be helpful in terminating Φ. But then agents sort through the index cards their hands picked out, and if they don't suffice to terminate Φ, they go through the box of thought content, as if one were going through a library index card file system, and pick out more index cards that might be helpful for addressing Φ. In the process, agents might accidentally grab one index card too many, or an index card of a different color, but no matter. Part of the skill involved in terminating Φ is setting aside the index cards that are not helpful but keeping the rest. Furthermore, agents stare at and attempt to relate index cards they have pulled out and determined to be helpful in terminating Φ to see if they can be connected in new and interesting ways that might help them terminate Φ. If they are, agents might create new index cards and compare them with other unified and non-unified index cards with the purpose of adhering

to their intention to terminate Φ. In any case, if agents repeat this process of retrieving, sorting, examining, refining, and unifying discrete contents with new connections, they intentionally terminate Φ.

With all this in mind, it is clear that Strawson wrongly claims that reasoning to a judgment cannot be a matter of intentional action because agents can only catalyze the process but not meaningfully control the unfolding of Φ between R_1 and Ω. Intentionally reasoning to a judgment consists not merely of continuously attending to the mental matter until the termination of Φ, which is the manifestation of the agent's sustained intention to reach Ω, but also of interaction with thought content in accordance with one's sustained intention to form a truthful judgment that adheres to the laws of reason by "Taking Stock" and "Unifying." An agent's sustained attention and their interaction with thought content between R_1 and Ω are, properly understood, exhibitions of agential control throughout the unfolding of process Φ. Moreover, the rule boundedness of reason does not eliminate the possibility of meaningful agential control such that the agent is merely intentionally present. The laws of reason do not happen to the agent in such a way that they magically structure their thoughts independent of their will. True, the agent accepts that the laws of reasoning are immutable, but they express their agency in their decision to entertain, examine, unify, and sort thought content consistent with their sustained intention to terminate Φ. Thus, for the reasons stated above, Strawson's skepticism that reasoning cannot be understood as a matter of action is untenable.

V. RETHINKING THE CONTROL CRITERION

I have rebutted, at length, Strawson's claim that much of the process of reasoning toward a judgment is sub-intentional, that it is dominated by large-scale automaticity, such that any process Φ can never be an intentional action. To do so, I have incorporated Wu's insight that an agent's continuous and sustained attention throughout Φ is just a manifestation of their continuous and sustained intention to form a truthful judgment in accordance with the laws of reasoning. Moreover, I have argued against the claim that an agent's continuous attending is not indication of control but simply of their mere presence, since, after all, they are not free to control inferential transitions. I argued that through Unification and Taking Stock, agents interact with thought content in a manner consistent with their intention and the rules of reasoning, but not such that it precludes them from controlling the unfolding of Φ. However, with all this being said, an astute interlocutor can exploit our minimal assent concerning the automaticity after R_1 to make the following argument:

We have granted that immediately after an R_1, as a matter of biological necessity, content relevant to terminating Φ comes to mind, just like that, without the agent themselves choosing it. Think, for example, of the fluidity involved in conversation, or of instances when, once a simple question is understood, thought content relevant to addressing it rapidly comes to mind without agents themselves choosing it. Can it also not be the case that, simply as a matter of

being the sort of beings that must make judgments with our limited but efficient brains, content relevant to terminating Φ comes to mind without the agent themselves choosing it on one or many occasions *after* R_1, in addition to the one experienced *immediately after* R_1?

For example, without the agent's conscious attendance, the brain might, in a flash, give them some important insight that they had previously not been thinking about. Similarly, the mind might suddenly press the agent to consider two distinct pieces of thought content by forcing them upon the scope of their conscious consideration. In addition to the fact that thought impingement is a common phenomenological experience, a recent neuroscientific study seems to confirm the idea that brains can make sense of data and detect patterns in inputs without participants' conscious awareness[163]. That could be the explanation behind why the brain often produces relevant thought impingement. Given that, in addition to the automaticity immediately after R_1, instances of thought impingement useful to the termination of Φ are conceivable and scientifically sound, does it follow that agents still do not meaningfully control the unfolding Φ and thus, that when they reason toward judgments their reasoning cannot be understood as a matter of action?

Further instances of absence of control over the coming to mind of thought content do not rule out the possibility that agents still consciously consider, entertain, examine, unify, and sort through thought content in a manner consistent with their sustained intention to terminate Φ. The brain might deliver flashes of information or force us to consider thought content that we had not been thinking about, but the agent must still do the work of checking to see whether the piece of information delivered is useful and consistent with their continued intention to form a truthful judgment. A sensible and alluring middle ground is obvious: process Φ is constituted of instances where the mind presents the agent with content independently of their choosing and of instances where the agent attends to thought content in ways consistent with their intention to form an Ω. However, in attempting to take this sensible middle ground, Strawson would immediately remind us of his basic assumption that a process Φ with an indefinite amount of intentional R_n's can never be intentional so long it is also composed of instances where the agent does not control the unfolding of the process. The only real way to completely address Strawson's skepticism is to reconsider the criteria that define what it means to act intentionally and, moreover, point toward the gradability of intentionality and control.

Up until now, we have understood intentional action as a special type of event consisting of an agent's intentional initiation and control throughout the unfolding of the process. However, because of the present counterargument, we must seek to make precise the relevant notion of control to escape the skepticism. Suppose that intentional action still involves the initiation and control criteria, but that we understand control as Strawson wants us to— as something that must be completely continuous, through and through, for an

163 See Naddaf (2024).

event to rise to the status of an intentional act. As such, a truly intentional act admits no instances of automaticity. However, if we understand control in this way, if one were to ask the classic Wittgensteinian question, "What is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?"[164] one would come up quite empty.

One would be able to say that the movement meets the first criteria of an intentional act, as agents intentionally initiate the raising of an arm, but since they do not control, through and through, the minutia of the psychosomatic mechanisms in charge of their arm raising, they fail to meet the second criteria. Is the correct response to say that agents do not intentionally raise their arm because they fail to meet both conditions on intentional action, for the sake of adhering to a set of criteria? If it's not apparent that we are in quicksand, now is the time to point it out. Through and through control is a delusion. There will always be a lack of agential control in the parts that constitute the execution of a mental or bodily movement. This is true of walking, closing our eyes, grasping and interfacing with objects in the world, standing, moving our mouths to speak, and, as we've identified for the mental domain, remembering, imagining, and even reasoning to a judgment. A strict, Strawsonian understanding of control quickly leaves us in a world where intentional action, both physical and mental, disappears entirely. We structure our world and our moral concepts around the existence of intentional action. So, if "there is mental action in every sense in which there is bodily action"[165] we must abandon our strict understanding of control. Rather than requiring complete, through and through continuity, there must be room for, minimally, instances of automaticity. In doing so, we restore the category of intentional action as something existent within our reality. [166]

Another resolution lies in pointing out that, perhaps, intentionality and control over many processes Φ lie in a continuum. If we could plot a continuum with all possible processes Φ, ordering them in relation to how many instances of automaticity they possess, there will be clusters of Φ's with zero or very few instances of automaticity, some with more instances of automaticity than the former cluster, and perhaps another cluster of Φ's composed of many instances of automaticity. The not so easy question then becomes, "When does an action stop being intentional?" Notice that the cluster of Φ's composed of many, or all instances of automaticity, does not necessarily mean these Φ's are any less intentional than clusters of Φ's that are less automatic. They might, for example, be Φ's undertaken with tremendous skill and dexterity in reasoning such that continued practice makes them feel automatic. Here, much would depend on whether one believes that very skilled action undertaken by reasoners, athletes, or musicians, is an intentional action. Regardless, I don't mean to say that I have an answer to this very important and complicated question.[167] Surely, however,

---

164 Wittgenstein (1958)

165 Strawson (2003, p. 231)

166 The thought that concurrent control is untenable is not a new idea. Similar conclusions are produced under Kelley's discussions of non-basic action and Wu's lofty, cognitive scientific treatment of ballistic action. See Kelley (2024) and Wu (2013).

167 For a possible solution, see Kelley, "The Normative Function of Intentional Action" (forthcoming).

there is much to be appreciated in the fact that we need not helplessly fall into Strawson's mental action skepticism. As I've shown, reasoning toward a judgment can in fact be an intentional action.

BIBLIOGRAPHY

Alvarez, M. (2009). Acting Intentionally and Acting for a Reason. *Inquiry: An Interdisciplinary Journal of Philosophy*, 52(3), 293-305.

Davidson, D. (1963). Actions, Reasons, and Causes. *Journal of Philosophy*, 60, 685-700.

Davidson, D. (1971). Agency. In R. Binkley, R. Bronaugh, A. Marras (Eds.), *Agent, Action, and Reason* (pp.1-37), University of Toronto Press.

Davidson, D. (2001). Freedom to Act. In D. Davidson (Ed), *Essays on Actions and Events* (pp.63-81). Oxford University Press.

Frankfurt, H. G. (1978) The problem of action. *American Philosophical Quarterly*, 15(2), 157-162.

Hieronymi, P. (2011). Believing at will. *The Canadian Journal of Philosophy*, 35, 149-187.

Kelley, M. (2024). How to perform a nonbasic action. *Nous*, 58(1), 106-125.

Kelley, M. (forthcoming). The normative function of intentional action. *Philosophers' Imprint*.

McKenna, M. (2022). Reasons-Responsiveness, Frankfurt Examples, and the Free Will Ability. In D. K. Nelkin & D. Pereboom (Eds.), *The Oxford Handbook of Moral Responsibility* (pp. 27-52), Oxford University Press.

Naddaf. M (2024). *How your brain detects patterns in the everyday: Without conscious thought*. Nature News. https://www.nature.com/articles/d41586-024-03116-8.

Schlosser, M. (2019). Agency. In Edward N. Zalta (Ed), *Stanford Encyclopedia of Philosophy*, Stanford University. https://plato.stanford.edu/archives/win2019/entries/agency/.

Setiya, K. (2008). Believing at will. *Midwest Studies in Philosophy*. 32, 36-52.

Stout, R. (2018). Ballistic Action. In R. Stout (Ed), *Process, Action, and Experience* (pp.210-228). Oxford University Press.

Strawson, G (2003). Mental ballistics or the involuntariness of spontaneity. *Proceedings of the Aristotelian Society*, 103(3), 227-256.

Vermaire, M. (2021). In search of doxastic involuntarism. *Philosophical Studies*, 179(2), 615-631.

Wittgenstein, L. (1958). *Philosophical Investigations*. Basil Blackwell.

Wu, W. (2013). Mental action and the threat of automaticity. In A. Clark, J. Kiverstein, & T. Vierkant (Eds.), *Decomposing the Will* (pp. 244-61). Oxford University Press.

Wu, W. (2022). Attending as mental action mental action as attending. In M. Brent & L. M. Titus (Eds.), *Mental Action and the Conscious Mind* (pp. 61-79). Routledge.

Uriel Vargas is a fourth-year philosophy student at the University of Chicago who doesn't like writing about himself in third person. Some of his recent achievements include getting lord Hulk on Marvel Rivals, reading 5 pages of Hegel in less than 5 hours (spectators say it was close!), and cutting a paper straight without looking like someone chewed the edges off. When he is not occasionally philosophizing, Uriel likes to watch movies, smell fragrances, eat ice cream, write satire and short stories, and say he goes to the gym.

AUTHOR'S NOTE

Many thanks are in order. I am tremendously indebted to Mikayla Kelley for her continuous support throughout the completion of this thesis project. Thank you for reminding me why I love philosophy. I would also like to thank Hannah McKeown for her enormously valuable feedback during our writing seminars, especially for pressing me on "rule-boundedness." I would also like to thank Camila Courreges for their never-ending curiosity and the immensely helpful feedback I received from the attendants at the B.A. Thesis Workshop and the editors of Logos. Thank you to my family and friends for the constant moral support any time I need it. Lastly, to my brother, Ricardo, my intellectual hero, who always encouraged me to ask questions and pursue my dreams, no amount of thanks can possibly repay the love and support you have given me throughout my life. If there is anything valuable in this piece of writing, know that it's because of you.

Freedom Through Necessity: A Defense of Spinozistic Compatibilism

Jinglong Yang
Vanderbilt University

FOREWORD

Editors: Marco Chong, Ethan Fisher-Dayn, Jesse Green, Amelia Klaus, Naveen Mathew, Hema Venkatraman

Colloquially, to be free is to be alive. Without the freedom to choose, life succumbs to a mere passenger ride. Maybe one feels the stomach drops of the ups and downs, but they possess little meaning because the outcome was always determined in advance. Consequently, when one begins to question whether we truly possess free will, deeper fears naturally follow. Can our choices matter? Can we change? Can we be responsible for what we do?

While many philosophers accept some form of determinism, perhaps there remains a middle position that preserves a meaningful sense of human freedom. In "Freedom Through Necessity," Jinglong Yang explores Baruch Spinoza's account that the world is determined by causality, yet one can still be free by being an adequate cause of their actions. The paper offers a careful reconstruction of Spinoza's position, challenges Matthew Kisner's influential interpretation that only God can be genuinely free, and responds to Leonardo Moauro's objection by distinguishing between acting from reason and acting from value judgments.

We valued this paper for its exceptionally close engagement with Spinoza's *Ethics*, and for its original contribution to the contemporary debate about freedom and determinism. In particular, Yang develops a nuanced interpretation of adequacy that preserves the possibility of human freedom while remaining faithful to Spinoza's strict necessitation framework. Readers who are pessimistic about the compatibility of determinism and a compelling form of human freedom may find Yang's vision to be shockingly optimistic. Simultaneously, this paper preserves an important ethical insight: even within a determined world, there remains a meaningful capacity to hold individuals accountable for their actions.

INTRODUCTION

Baruch Spinoza's philosophy presents one of the most uncompromising forms of determinism in the history of philosophy. According to him, all things follow from God's nature with the same necessity by which it follows from the nature of a triangle that its angles equal two right angles. Human beings, as finite modes, are no exception. Spinoza believes that our actions, thoughts, and desires are entirely determined by the causal order of nature. As a consequence, his determinism is in direct conflict with libertarian conceptions of free will—the notion of self-caused choice between alternatives. With a rejection of such free will, Spinoza's thorough determinism seems to generate a problem: if all human actions are necessary and determined in advance, the foundation for moral responsibility seems to collapse. Thus, a challenge naturally arises: Can Spinoza have an account of human freedom without presupposing free will that

is compatible with his determinism? I believe the answer is "Yes!" However, it requires a radical revision of the idea of freedom.

In this paper, I will delve deeply into Spinoza's unusual compatibilist vision. First, I will reconstruct Spinoza's determinism, showing how his metaphysics strictly rejects free will and contingency. Next, I will construct a positive account of human freedom according to my interpretation of Spinoza's *Ethics*. Spinozian scholars have commonly identified causal independency—the ability to be an *adequate cause* of one's action—as the source of freedom (Rocca, Sleigh, and Chappell, 1988, pp. 1230-1231). Whether anything besides God is capable of acquiring such independence has been a matter of debate. Matthew Kisner (2011; 2021) has famously argued that humans cannot acquire such causal independence, because we cannot be "adequate causes" of our actions. I will construct my positive account of freedom primarily by arguing against Kisner's position. Then, I will demonstrate how this account of freedom is compatible with Spinoza's determinism, resolving the apparent contradiction between being determined and being free. In this final section, I will consider a powerful rejection raised by Leonardo Moauro (2023), who argues that Spinoza's theory resists a compatibilist reading of free agency due to "value projectivism." I will respond to Moauro's critique by showing that Spinozian compatibilism is radically different from any standard account of compatibilism, thereby escaping the powerful critique he charges.

Together, these arguments demonstrate that Spinoza offers an unusual but coherent form of compatibilism in which human freedom consists not in exemption from causal determination but in obedience to it. If Spinoza is correct about determinism, then our minds and bodies are fully determined, whether we realize it or not. However, we can become "freer" through adequate understanding of oneself and of Nature (or, in Spinoza's language, God), i.e. when we align our desires and actions with our nature and with the unfolding of the universe. Essentially, Spinoza's conception of freedom is not undermined by necessity but made possible by it.

SPINOZA'S DETERMINISM IN THE *ETHICS*

First, let us begin our discussion with Spinoza's uncompromising determinism. According to Spinoza's metaphysics, everything that exists and happens follows from the necessity of Nature (or God). There is no room for contingency or chance. He states this explicitly in E1p29:

"Nothing in nature is contingent but everything is determined to exist and to operate in a specific way by the necessity of the divine nature."

In other words, the entire universe is an interconnected chain of *causes*. Every finite thing is caused by another finite thing, and by another, and so on to infinity. This infinite chain of causes and effects is ultimately rooted in God's eternal essence. Whatever occurs could not have been otherwise, given the total order of Nature.

Crucially, Spinoza's determinism extends to human actions, thoughts, and

decisions. In E2p48, Spinoza explicitly denies the existence of free will in the mind:

"There is no absolute or free will in the mind, but the mind is determined to will this or that by a cause which is also determined by another cause, and this in turn by another, and so on ad infinitum."

Clearly, for Spinoza, our so-called choices are not exceptions to the causal order. If we had complete knowledge of any decision a person makes, we could trace all of its determining causes—such as prior psychological states, external influences, the individual's innate temperament, and so on—ultimately back to God or Nature. Thus, Spinoza's metaphysics aligns with strict incompatibilism regarding libertarian free will. Given that the world is completely determined, the will cannot be free in the sense of an uncaused, independent power (E1p32c1, E2p48s).

One crucial piece of Spinoza's determinism is his psychological error theory, which explains why people falsely believe they possess free will. In E2p35s, he writes:

"...human beings are mistaken in thinking they are free. This belief consists simply of their being conscious of their actions but ignorant of the causes by which they are determined."

This shows that we experience our decisions directly, but we typically do not perceive the myriad external and internal factors that influence them. A feeling of choosing freely may simply indicate that we do not perceive external compulsion, yet our internal motives have been shaped by many factors we did not choose. In *Letter 58*, Spinoza offers us a vivid analogy: if a falling stone became conscious, it would believe it was freely choosing to fall, since it is aware of its motion but not the boy who threw it. Hence, for Spinoza, the common notion of free will is just a product of human ignorance rather than an autonomous faculty.

It is important to note that Spinoza's denial of free will extends even to God. Contemporary scholars such as Don Garrett (2018) have rigorously demonstrated that Spinoza rejects the image of God as a volitional creator who could have willed things to be otherwise. As Garrett explains, God acts by the necessity of his own nature alone, not by free choice. In E1p17c2, Spinoza writes that only God is absolutely free in the sense that nothing external constrains Him, but even God does not "will" alternatives. Instead, everything inevitably follows from God's nature. On this view, Spinoza has firmly ruled out free will in the libertarian sense as a pure fiction.

HUMAN FREEDOM IN A DETERMINISTIC WORLD

Despite rejecting free will, Spinoza insists that humans can achieve a genuine form of freedom. In fact, he dedicates a major portion of Book IV and V to show us the "way leading to freedom." (E5pref) To understand what he takes genuine human freedom to be, we must shift our attention away from the will and its alleged power of choice and toward the role of reason, activity, and necessity.

In Book I of his *Ethics*, Spinoza explicitly defines a thing to be free if "it exists solely from the necessity of its own nature, and is determined to action by itself alone." In contrast, a thing is compelled (or necessary) if "it is determined by another thing to exist and operate in a specific and determinate way." (E1def7) Hence, freedom for Spinoza is not an exemption from causal determination but a distinctive way of being determined: namely, self-determination in alignment with one's nature. Under his absolute determinism, only God is free in the strictest sense, because God is the only being capable of complete causal independence—acting from his nature alone, undisturbed by any external factors. After all, nothing is external to God; everything is within God. Thus, God necessarily causes Himself.

However, freedom is a spectrum; it comes in degrees. Spinoza does acknowledge that a certain degree of freedom can be attained by humans. Although the freedom available to us may be partial and imperfect, the fact that human freedom is possible would save Spinoza's determinism from being accused of dismissing moral responsibility and undermining the foundation of ethics and moral philosophy. Whether the source of human freedom is causal independence has been a major point of debate.

Matthew Kisner argues that Spinoza's metaphysics strictly rule out the possibility of human freedom understood as causal independence. Given the strict definition of the term "free" in E1def7 and the uncompromising determinism I have demonstrated in the last section, Kisner believes that no human being can be an adequate cause of his own action.

The heart of Kisner's argument lies in his identification of "adequate cause" with the "sole cause." Spinoza defines an adequate cause as one "whose effect can be clearly and distinctly perceived through the said cause," and contrasts it with an "inadequate or partial cause" whose effect "cannot be understood through it the said cause alone" (E3def1). Kisner's key move is to treat "understand through the said cause" as a no-remainder condition: if anything besides the said cause is required for understanding the effect, then the cause is, by definition, only partial and inadequate. Kisner then appeals to an important insight from Spinoza: to understand an effect is to understand its causal antecedents. Given this premise, he infers that if an effect can be understood "clearly and distinctly" through some purported cause, then there cannot be any other causal antecedents that would also need to be understood; therefore, the purported cause must be the only causal antecedent of the effect. This is why he concludes that "the definition (Spinoza's definition of adequate cause) stipulates that an adequate cause is the only or sole causal antecedent of an effect." (Kisner, 2011, pp. 25-27)

Kisner then leverages this sole-cause interpretation inside a broader argument about adequate ideas. He claims (i) adequate ideas must represent a thing with its causal antecedents, and (ii) because those antecedents are the thing's adequate cause (i.e., its sole cause), having an adequate idea of a thing entails being the sole cause of that idea (pp. 27-28).

An important consequence of Kisner's argument is that it implies humans cannot conceive adequate ideas. E1p17c2 notes that "God alone is a free cause",

because only God can "exist by the necessity of his own nature, and acts solely by the necessity of his own nature." Kisner believes this implies that humans cannot be causally independent in determining their actions and, consequently, cannot be adequate causes (p. 46). Since being an adequate cause is coextensive with conceiving adequate ideas for Kisner, it follows that humans cannot conceive adequate ideas either. The only entity that can do so is God.

This is a serious problem, because Spinoza explicitly claims that humans *can* conceive adequate ideas (E2p38c, E2p47), and doing so is a central part of his *Ethics*. Kisner acknowledges this tension between the consequence of his argument and Spinoza's claim, but does not consider it seriously. He tries to resolve the tension by charging Spinoza with equivocation when using the terms "adequate cause," when referring to God, and "adequate idea," and with humans. Humans, Kisner argues, can only be adequate causes and have adequate ideas in a weaker sense of the term (pp. 46-47). In other words, humans can only approximate having adequate ideas, but never truly attain them.

However, Kisner's attempt to resolve this tension by appealing to a "weaker" sense of adequacy ultimately fails, because it abandons the very notion of adequacy his argument requires. Recall that Kisner treats the definition of adequate causes as a no-remainder condition: if an effect cannot be understood through the cause alone, then the cause is, by definition, only partial. This leaves no room for anything to be a partially adequate cause. Any adequate cause must be the sole cause of its effect. Accordingly, any partiality necessarily entails inadequacy.

Since Kisner claims that being an adequate cause is coextensive with conceiving adequate ideas, the same no-remainder standard transfers to adequate ideas as well. In other words, humans can only have adequate ideas if they are produced through the human mind alone, without requiring anything else for their causal explanation. Once this argument is granted, to say that humans possess adequate ideas only "in a weaker sense" is simply to say that humans possess them partially. However, partial understanding is precisely what Spinoza calls inadequate understanding. Thus, Kisner's "weaker adequacy" collapses into inadequacy, and humans can at best form inadequate ideas. The original tension remains intact—Spinoza's explicit claim that humans can conceive adequate ideas is not explained, but effectively denied.

The persistence of this tension, I suggest, originates from a misreading of Spinoza's definition of adequate cause. Spinoza defines an adequate cause as one "whose effect can be clearly and distinctly perceived through the said cause" (E3def1). Kisner's interpretation places decisive weight on the phrase "through the said cause," treating it as a no-remainder condition that forces adequacy to entail sole causation. However, a more natural emphasis, and the one that best fits Spinoza's broader epistemology, falls on the phrase "clearly and distinctly perceived." I believe that the point of this definition is not to impose metaphysical exclusivity, but to mark a standard of intelligibility: a cause is adequate insofar as it provides an explanation sufficient for a clear and distinct understanding of the effect.

This reading is supported by Spinoza's wording of E3def1. Spinoza does not define the adequate cause as that through which an effect is understood "through the cause alone"; the word "alone" only appears in the definition of inadequate ideas: "I call it partial, or inadequate, if its effect cannot be understood through it alone." This asymmetry matters. It shows that Kisner's inference from "through the cause" to "the sole cause" requires inserting a restriction that Spinoza does not state.

Once the emphasis is correctly placed on the "clear and distinct" clause, it becomes permissible for humans to conceive adequate ideas. According to E1p28, every finite mode—including every action and thought—is determined by an infinite series of prior causes. If Kisner's sole-cause interpretation were correct, this would render human adequacy impossible from the start, since no finite cause could ever exhaust the totality of causal antecedents. Yet Spinoza explicitly affirms that humans can form adequate ideas of certain things (E2p38s, E2p47). The resolution lies in recognizing that not all causal antecedents are explanatorily relevant to producing a clear and distinct understanding of an effect. The further an antecedent cause is located in the causal chain, the less explanatory power it contributes to understanding the effect. Adequacy, then, does not demand an exhaustive inventory of every prior cause; it demands only those causes sufficient to render the effect intelligible. Being an adequate cause, on this reading, does not entail being the sole cause.

Since adequate causes and adequate ideas are coextensive, the same standard of explanatory sufficiency applies to adequate ideas. Spinoza defines an adequate idea as one possessing all the 'intrinsic characteristics of a true idea" (E2def4). In his explanation, he stresses that these intrinsic characteristics are distinct from the traditional correspondence criterion.

Spinoza does hold that a true idea "agrees" with its object (that is the traditional correspondence formula) (E1ax6), but he treats this agreement as an extrinsic characteristic: a relational feature connecting *this* idea to *that* thing. The intrinsic characteristics, by contrast, refer to features internal to the idea itself that render it intelligible. An idea is adequate when it is internally articulated in such a way that the mind can grasp why the content must be so. Such intelligibility does not require representing the complete causal history of an object; it requires only that the idea captures those features necessary for clear and distinct understanding.

To illustrate Spinoza's conception of adequate ideas, consider the following example. An uneducated man may form the inadequate idea that "the sun is about 200 feet away." This is a confused idea caused by how his body is affected by light and perspective. The idea may reliably arise under certain conditions, but his mind does not grasp why the sun appears that way. His idea of the sun's distance is confused because he possesses the representation without understanding its grounds. By contrast, an adequate idea of the sun's distance would be grounded in geometry and physics. With proper knowledge, one can see why the distance must be what it is. However, this adequate idea needs only to be connected to other relevant adequate ideas in a rational chain—for

instance, the geometric principles of parallax and triangulation (the distance to a remote object can be calculated by measuring angles from two known positions) and Kepler's third law, which relates orbital periods to distances from the sun. The idea need not include every factor in the causal history of how that understanding was formed, such as a complete derivation of Euclid's geometrical theorems from first principles or a full account of how Kepler arrived at his laws. In this respect, an adequate idea resembles a geometric demonstration more than a genealogical record. What matters is the internal coherence and sufficiency of the explanation, not its causal completeness.

With this interpretation in place, Kisner's charge that Spinoza equivocates on "adequate cause" and "adequate idea" dissolves. The apparent tension arises only if adequacy is interpreted as metaphysical exclusivity—that is, as requiring sole causation. Once we recognize that adequacy denotes epistemic sufficiency instead, the gap between divine and human adequacy becomes one of scope rather than kind. God, as the infinite substance, is the adequate cause of all things; humans, as finite modes, can be adequate causes of some things. In both cases, the standard remains the same: a cause is adequate when it suffices for a clear and distinct understanding of its effect. Thus, both God and humans can conceive adequate ideas.

The ability to conceive adequate ideas is a crucial function of reason for Spinoza. Establishing that humans can conceive adequate ideas not only preserves the theoretical coherence of his philosophy but also illuminates the relationship between reason and freedom: the more adequate ideas we acquire about ourselves, the freer we become. To better understand why this is, we need to draw insights from how Stuart Hampshire shows that adequate ideas have causal power.

Hampshire (1971) identifies a distinctive feature of psychic causality that grounds the causal efficacy of adequate ideas. As he points out, mental causation operates reflexively: when someone reflects on the causes of his beliefs, desires, or emotions, the very act of reflection directly modifies those mental states (p. 560). Hampshire emphasizes that thoughts are caused by other thoughts, and the laws governing their concatenation are either laws of association or laws of rational necessity (p. 561). Most of our beliefs and desires are formed passively through association, without conscious reflection. However, when we actively inquire into the causes of our mental states—that is, when we form adequate ideas of them—we replace associative links with rational connections, thus transforming the original states. As Hampshire puts it, if a man's beliefs about his own states of mind are changed through reflection, his emotions, attitudes, and consequently his conduct change as well (p. 565). Adequate ideas thus possess genuine causal power: they do not merely represent the mind's condition but actively restructure it.

The acquisition of adequate ideas thereby increases our freedom. The more adequate ideas we acquire about ourselves, the more we are able to align our desires with rational desires, thereby acting from the nature of who we truly are. And the more adequate ideas we acquire about this world, the more we are able

to see through false associations between events and discern the true connection behind them. Both of these effects of acquiring adequate ideas would increase our *conatus* (power of acting) and make us more of an adequate cause of our own actions, rather than being acted upon and compelled by external causes. Essentially, the more adequately rational one is, the more he is able to act from his own nature; and the more one acts from his nature, the freer he becomes.

RECONCILING FREEDOM AND DETERMINISM IN SPINOZA'S SYSTEM

At this point, we have completed two pieces of the puzzle: (i) determinism—every event, including human actions, is determined by prior causes, and free will is an illusion; (ii) freedom—a person can be free if they act from their own nature. Now we must show clearly how these two pieces fit together. In other words, why doesn't Spinoza's determinism undermine his account of freedom? What makes his view genuine compatibilism rather than a contradiction?

The key lies in the redefinition of freedom stated at the end of the last section. The classical tension between free will and determinism assumes that freedom means the ability to do otherwise. Under that definition, if determinism is true, then indeed no one can do otherwise and hence no one is free. Spinoza sidesteps this dilemma by denying that freedom has anything to do with the contingency of choice (E1p32c1, E2p48). The availability of alternative possibilities is not what makes us free. In fact, such alternatives are illusory, since only one outcome can occur given prior causes. Instead, Spinoza locates freedom in the source of our actions. If the source is internal—if our actions spring from our own nature—then the action is free. If the source is external—if something outside us compels or triggers the action—then the action is unfree. This shift in perspective makes freedom a matter of how an action is caused, not whether it is caused at all. It is this redefinition of freedom that makes Spinoza an unusual compatibilist.

If the key to freedom lies in acting from our own nature, then we must ask: what is human nature for Spinoza? One answer is that we are, at our core, rational animals. Throughout Books IV and V of the *Ethics*, Spinoza frequently identifies reason as intrinsic to human essence. In E4p35c1, he explicitly claims that "a man acts by the laws of his own nature absolutely when he lives by the command of reason." Similarly, in E4p26, he claims that whatever we strive for from reason is nothing but adequate understanding; when uncorrupted by external influences, the mind's deepest drive is always toward rational comprehension. This explains why Spinoza ties freedom so closely to rationality. To act from reason is to act from adequate ideas, and to act from adequate ideas is to be an adequate cause of one's actions. Reason, in other words, is the faculty through which we become self-determined rather than externally driven. The passions pull us in conflicting directions according to the winds of fortune; reason unifies our striving and directs it toward what genuinely enhances our power, our *conatus*. Thus, Spinoza presents the "free man" as a human ideal toward which we should all strive. "The free man always acts honestly, not

deceptively" (E4p72) because he sees the truth clearly, understands the necessity of the order of Nature and aligns himself with the unfolding of reality.

This alignment of oneself to the order of Nature is central to understanding Spinoza's compatibilism. When we adequately understand something, we recognize why it had to be so. The "free man" who lives by the "dictates of reason" consents to the unfolding of Nature precisely because he comprehends its necessity. He does not futilely rail against what cannot be otherwise; instead, he aligns his desires with the way things must be. Thus, Spinoza's vision for the free man is not someone who can act differently, but one who acts indifferently in accordance with the order of Nature. We are most free when we assent to Nature's necessity because we see the truth clearly. To put it simply: to be free is to understand necessity, and to understand necessity is to say yes to it.

Finally, this reconciliation of freedom and determinism carries an ethical implication: it preserves autonomy and responsibility without invoking a contra-causal free will. Spinoza can coherently maintain that some people are freer—and, in a sense, more responsible for their actions—than others, not because they could have magically chosen otherwise, but because they have developed their rational capacities and thereby become the adequate causes of their own actions. On this view, moral praise and blame transform into an evaluation of one's adequacy to one's own rational self-government. We don't blame someone as though they could have done otherwise, but we can still distinguish the slave of passion from the man of reason and prefer the latter as a model. In this way, Spinoza retains a notion of autonomy and agency, despite his strict determinism. A rational agent is not an uncaused causer, but a relatively self-caused actor: one whose actions flow from adequate ideas and are thus adequately explained by reference to one's own nature.

All these suggest that freedom and determinism, for Spinoza, do not stand at odds against each other; instead, they are peacefully coherent. The more we understand the necessary order of things, the more our actions arise from that understanding and the freer we become.

However, compatibilist readings of Spinoza usually face a serious challenge from Leonardo Moauro's interpretation of Spinoza's moral psychology. On a standard compatibilist account, freedom requires not merely self-determination but intelligent self-determination: we act freely when our actions flow from our judgements about what is good. Yet Moauro contends that Spinoza inverts this order of explanation. In the *Ethics*, Spinoza claims that:

"We do not endeavor anything, we do not will anything, we do not seek or desire anything, because we judge it to be good; on the contrary, we judge a thing to be good because we endeavor it, will it, seek it and desire it." (E3p9s)

Thus, our value judgments do not cause our desires; rather, our desires cause our value judgments. Moauro names this "value projectivism": the mind projects the property of goodness onto objects based on its appetites, then mistakenly experiences these projected values as the causes of its actions (Moauro, 2023, pp. 9-12). On this reading, good and evil are simply "beings of imagination"—not real properties of things but confused representations arising from how

objects affect us (E1app). If Moauro is correct, then people who act based on what they think is good are not free at all. Their actions are determined entirely by appetites that are blind to value judgments, and their sense of acting for reasons is just an error built into the structure of human cognition.

Although Moauro has constructed a powerful critique, his objection applies only to the standard account of compatibilism, not to the Spinozistic version I have proposed. Standard compatibilism, exemplified by Leibniz, typically requires three conditions for freedom: (i) spontaneity (self-determination), (ii) contingency (actions are not metaphysically necessary), and (iii) intelligence (actions are determined by representations of value) (Moauro, 2023, p. 3). My reading of Spinoza denies conditions (i) and (ii) outright. On the standard interpretation of Spinoza's metaphysics, only God can be entirely self-determining, and God's nature admits no contingency in the order of things. As I have shown in section two, my proposed compatibilism fully accepts this standard reading. Indeed, the distinctive value of a Spinozistic compatibilism lies precisely in its capacity to reconcile human freedom with partial dependence on external causes and complete causal necessity—a reconciliation I have demonstrated above. But there is a further difference between my account and the standard one: on my reading, to act from reason is not to act from value judgments.

As I have argued, acting from reason, for Spinoza, should be understood as acting from adequate ideas. To recall, Spinoza defines an idea to be adequate if it contains all the "intrinsic characteristics of a true idea." (E2def4) Since moral values like good and evil are merely "beings of imagination" but not real properties of things, they cannot qualify as intrinsic characteristics that correspond to the representational content of any object. In fact, moral values do not belong to objects at all—they are human constructs projected onto things based on our appetites. Hence, to act from value judgments is to act from confused ideas, not adequate ideas, and thus to fail to act according to reason. However, freedom, for Spinoza, requires precisely that we act in accordance with "the dictate of reason" (E4p67). The free man is not one who acts on what he imagines to be good; he is one who acts from what he adequately understands.

My account of Spinozistic compatibilism thus escapes Moauro's critique by rejecting the very framework his objection presupposes. Moauro's value projectivism undermines compatibilist views that locate freedom in the capacity to act on representations of value—but my reading does not locate freedom there. On my account, freedom does not consist in being moved by value judgments; it consists in acting from adequate ideas, which are explanatorily sufficient for a clear and distinct understanding of one's actions. Value judgments, being products of the imagination, represent precisely the kind of confused cognition from which the free man liberates himself. Freedom, on this reading, is not threatened by the fact that good and evil are projections of appetite. On the contrary, recognizing this fact is part of what it means to think adequately. The free man acts from reason, not because he is guided by imagined goods, but because he understands the necessary order of nature and aligns his striving with that understanding.

CONCLUSION

Now, I will conclude my defense for a compatibilist reading of Spinoza's *Ethics*. At this point, two significant challenges to the coherence of human freedom within Spinoza's deterministic worldview have been addressed. Against Kisner's claim that adequate causation requires sole causation—and thus lies beyond the reach of finite beings—I have argued that Spinoza's definition of adequate cause emphasizes epistemic sufficiency rather than metaphysical exclusivity. The key phrase in E3def1 is "clearly and distinctly perceived," not "through the said cause alone." Once this interpretive shift is made, humans can genuinely conceive adequate ideas and be adequate causes of their actions. Against Moauro's challenge that value projectivism undermines rational self-determination, I have shown that Spinozistic freedom does not consist in acting on value judgments—which are indeed products of the imagination—but in acting from adequate ideas. The free man is not guided by imagined goods, but by what he adequately understands.

The significance of this research extends beyond Spinoza scholarship. In contemporary philosophy of action, compatibilism remains a contested position. Critics often charge that compatibilist accounts of freedom are merely nominal—that calling someone "free" while admitting their actions are determined is an empty gesture. Spinoza offers a robust response to this worry, using a radical redefinition of freedom. Freedom, in his view, is not a metaphysical exemption from causation but a distinctive mode of being caused: namely, being determined by one's own nature rather than by external forces. We become freer not by escaping the causal order of the universe but by comprehending it and finding our place within it.

As a result, Spinoza's compatibilism points toward a distinctive form of liberation. Once we truly align our actions not only with the necessity of our own nature but also with the necessity of God's nature, we are freed from all compulsion. In true liberation, we will allow things to come and go as they do, without developing any aversion to their changes. Such a peace of mind is the kind of "blessedness" Spinoza ascribes to the "free man." It is as though one entrusts oneself entirely to the stream of activity in the universe and flows with it. When properly understood, we will see that this deterministic world is not a prison in which the free man is all chained up by necessity; instead, it is a symphony in which harmony resounds through all things, and the free man is an inseparable part of it. After all, to be free is to understand necessity, and to understand necessity is to say "yes" to it!

BIBLIOGRAPHY

Garrett, D. (2018). Spinoza's Necessitarianism. In Necessity and Nature in Spinoza's Philosophy

(pp. 98-124). Oxford University Press.

Hampshire, S. (1971). Spinoza's Theory of Human Freedom. The Monist, 55(4), 554-566. Kisner, M. (2021). Freedom without Independence. In N. Naaman-Zauderer (Ed.), Freedom,

Action, and Motivation in Spinoza's Ethics (pp. 37-61). Routledge.

Kisner, M. J. (2011). Freedom as Rationality. In Spinoza on Human Freedom: Reason, Autonomy and the Good Life (pp. 17-45). Cambridge University Press.

Moauro, L. (2023). Spinoza on Freedom, Feeling Free, and Acting for the Good. Argumenta, 1, 1-16. DOI 10.14275/2465-2334/20230.moa

Spinoza, B. (2017). Correspondence. Early Modern Text. https://www.earlymoderntexts.com/assets/pdfs/spinoza1661.pdf

Spinoza, B. d. (2018). Ethics, Proved in Geometrical Order (M. J. Kisner, Ed.; M. Silverthorne & M. J. Kisner, Trans.). Cambridge University Press. 10.1017/9781107706972

Sleigh, Robert Jr., Chappell, Vere, and Della Rocca, Michael. (1998). "Determinism and Human Freedom," in D. Garber and M. Ayers (eds.), The Cambridge History of

Seventeenth-Century Philosophy, vol. 2, Cambridge: Cambridge University Press, 1195–1278.

Jinglong Yang is a fourth-year undergraduate student at Vanderbilt University, studying Philosophy, Economics, and Classical Voice Performance. His philosophical interests include studies on ethics—in particular, the normative competence of AI agents—and metaphysics. He is also interested in Hellenistic philosophy, moral philosophy, and potential for the integration of ancient ways of thinking into modern society. In an age of rapid technological advancement, philosophy is urgently called for to guide and check the increasing power of science. His research is an endeavor in this direction. In the hopes of becoming a lifelong academic, he plans to pursue a PhD in Philosophy.

AUTHOR'S NOTE

Spinoza has gradually become my favorite philosopher in the early Modern period. His unusual notion of God, radical revision of freedom, and the concept of conatus have given me valuable tools for navigating what turned out to be one of the most difficult periods of my life. Writing this paper on freedom has been particularly enjoyable. It opened up my interest in not just metaphysics, but also philosophy of mind, psychology, and moral philosophy. I was impressed by how versatile Spinoza's philosophy can be and amazed by how his seemingly obscure philosophy can play a practical role in helping me live a better life.

Importantly, I'd like to take a moment and thank Professor Emanuele Costa for introducing Spinoza into my life. If not for his teachings and insights, I probably would not have benefited so much from Spinoza's philosophy, both personally and intellectually.

Also, I'd like to thank anyone who has spent the effort reading this article of mine. If Spinoza is right about determinism, perhaps it is fate that we meet each other in thoughts in this special manner. I hope you have enjoyed reading this paper, just as I have enjoyed writing it. If you also happen to be amazed by Spinoza and would like to talk, please don't hesitate to reach out. I'd love to connect!

The Simplest Solution to the Deepest Paradox of Deontic Logic is Not so Simple After All

Diana Nichvoloda
Harvard College

FOREWORD

Editors: Alexander Coulouras, Julian Kanu, Robert Wallace Sims, Emily Zhou

Deontic logic is the logic of obligations and permissions. It is a variant of modal logic that interprets the standard necessity operator □ as "it is obligatory that" and the possibility operator ◇ as "it is permissible that." In other words, for a proposition p, Op reads as "it is obligatory that p," and P reads as "it is permissible that p." Thus, the formulas of deontic logic are generated by a set of propositional constants S together with the connectives {∨ (or), & (and), → (if–then), ~ (not), O, P}.

Given seemingly intuitive axioms — O(p→q)→(p→Oq), the standard obligation form of the K axiom (p→q)→(Op→Oq), and Op→~O~p — one can derive an absurd conclusion. We also assume two intuitive normative statements: (1) a person named Smith ought not to kill Jones, and (2) it ought to be the case that if Smith kills Jones, then he does so gently. Statement (1) is straightforward, and statement (2) reflects the idea that, if someone is killed, the best-case scenario is that it is done as gently as possible. Together, these assumptions allow us to derive a formal contradiction in deontic logic. This is known as the Gentle Murder Paradox (GMP).

One response to this problem, offered by Joaquin Jeremiah Joven, claims that the GMP is ill-formed in deontic logic. The paradox depends on applying the obligation operator O to indicative statements, whereas, according to Joven, it should apply only to imperatives. This paper responds by developing a more fully articulated formal system based on Joven's commitments and then arguing that, within this revised system, the original paradox can be reformulated. Therefore, Joven fails to provide a compelling solution to the GMP.

Our main choice to publish this paper rested on Diana Nichvoloda's impressive use of techniques in formal logic to forward the debate around the GMP. In addition, Nichvoloda deftly engages with the contemporary literature on deontic logic. Our readers who are interested in formal linguistics and philosophical logic may find value in reading this paper.

A version of this paper is also published in *Aurantium*, Syracuse University's undergraduate philosophy journal.

INTRODUCTION

In his 2023 paper, *The Simplest Solution to the Deepest Paradox of Deontic Logic*, Jeremiah Joven Joaquin argues that the problem with the Gentle Murderer Paradox, one of the deepest paradoxes in deontic logic, arises as a result of an incorrect formulation of the paradox, in which the deontic operator denoting obligation is applied to indicative statements.[168] His proposed "simplest solution"

168 Jeremiah Joven Joaquin, "The Simplest Solution to the Deepest Paradox of Deontic Logic," *Organon* F 30, no. 4 (2023): 312–322.

is to restrict its scope to only imperative statements. Joaquin also briefly considers and rejects an objection, which would revise one of the axioms of deontic logic, as applied to the paradox, so that it is stated only in terms of imperatives. In this paper, I first give a general introduction to deontic logic, explain its notation and axioms, and then apply it to further regiment the Gentle Murderer Paradox. I then consider the objection raised in Joaquin's paper, and show that if Joaquin's reply to it is properly fleshed out by restating *all* of the deontic principles in terms of imperatives, then applying them to the Gentle Murderer Paradox, the paradox still stands, so Joaquin has not, in fact, found a solution to it.

THE GENTLE MURDERER PARADOX

I first outline the Gentle Murderer Paradox as first presented by James William Forrester in his 1984 paper, and as restated by Joaquin. Forrester begins with two reasonable assumptions about murder:

(A1) All kinds of murder are forbidden.

(A2) In a case where murder is unavoidable, murdering someone violently is worse than murdering someone gently.

Forrester and Joaquin then make the following assumption: in fact, Smith did murder Jones. The paradox is stated as follows:

1. It is obligatory that Smith does not murder Jones.
2. It is obligatory that if Smith murders Jones, Smith murders Jones gently.
3. If Smith murders Jones, it is obligatory that Smith murders Jones gently.
4. Smith murders Jones.
5. It is obligatory that Smith murders Jones gently.
6. But if Smith murders Jones gently, then Smith murders Jones.
7. It is obligatory that Smith murders Jones.
8. It is and it is not obligatory that Smith murders Jones.

As a first step to understanding why this paradox arises, Joaquin regiments it using the axioms of propositional and deontic logic and defines clearly how he arrives at each of the steps and the conclusion.

THE PRINCIPLES OF DEONTIC LOGIC

Modal logic extends classical propositional logic by introducing operators representing necessity and possibility. Just as in classical logic, formulas are constructed from the usual propositional variables (such as *p, q, r*) using the usual propositional connectives: ¬ (negation), . (conjunction), ∨ (disjunction), ⊃ (material conditional), and ⇔ (biconditional). The symbol ⇒ is used to represent entailment. However, where classical logic concerns itself only with truth values at a single world, modal logic focuses on a set of possible worlds and the relations that hold among them. Instead of evaluating a claim solely as true or false, evaluating a claim at a certain world a involves examining whether it holds in every world accessible from that world, or whether it holds in at least one such world. To formalize this, modal logic introduces two additional operators:

- □p, which means "p is necessary,"
- ◇p, meaning "p is possible."

In classical systems of modal logic, possibility can be defined in terms of necessity: $\Diamond p \equiv \neg\Box\neg p$. This means that something is possible if its negation is not necessary. Modal logic interprets statements involving □ and ◇ using models based on possible worlds. A model specifies a set of worlds and an accessibility relation between them, allowing us to evaluate whether a formula is true at a given world. This framework can be used to represent different kinds of modality, including metaphysical possibility, knowledge, belief, and obligation.

Deontic logic specifically applies the modal framework to questions of obligation and permission. Obligation, the idea that a certain action, state of affairs, or proposition *ought* to be the case, is the deontic equivalent of necessity, and permission is the equivalent of possibility. Thus, the deontic operators are defined as follows:

- O*p*, which means "it is obligatory that *p*,"
- P*p*, which means "it is permissible that *p*."

Just as ◇p is classically equivalent to $\neg\Box\neg p$, the deontic operator P is typically defined as:

$P\boldsymbol{p} \equiv \neg O\neg p$.

Put simply, this means that if not-*p* is not obligatory, then *p* is permissible.

Deontic logic has three core principles, which Joaquin abbreviates to D1, D2, and D3.

The first is the principle of conditional obligations:

(D1) $O(\boldsymbol{p} \supset \boldsymbol{q}) \supset (p \supset Oq)$,

meaning that if *q* is obligatory given *p*, then, should *p* obtain, it follows that *q* is obligatory. The second is the inferential rule:

(D2) $(\boldsymbol{p} \supset \boldsymbol{q}) \Rightarrow (O\boldsymbol{p} \supset Oq)$,

which somewhat controversially allows us to infer $O\boldsymbol{p} \supset Oq$ from the material conditional $\boldsymbol{p} \supset \boldsymbol{q}$. This means that whenever the material conditional $\boldsymbol{p} \supset \boldsymbol{q}$ is true—even in cases when it is vacuously true—the corresponding conditional about obligations follows. Thus, D2 is a way to show how material conditionals transfer into deontic conditionals.

The third principle is

(D3) $O\boldsymbol{p} \supset \neg O\neg p$,

which states that if *p* is obligatory, then it is not obligatory that not-*p*, i.e., that whatever is obligatory is permissible.

REGIMENTING THE PARADOX

In this section, I explain how Joaquin uses deontic logic to explain Forrester's paradox and how this leads him to the conclusion that the problem with the paradox is its formulation, specifically the scope of the O operator. Though in his paper, Joaquin doesn't regiment the Gentle Murderer paradox only in terms of propositional variables and connectives, I think it is important to do so in order to see its structure and the logic of deriving each step sequentially more clearly.

Let *a* represent "Smith murders Jones gently" and *b* represent "Smith murders Jones." From the two assumptions (A1) and (A2), we have

Step 1. O(¬*b*) Step 2. O(*b* ⊃ *a*)

From Step 2, using D1:

Step 3. *b* ⊃ O*a*

By assumption:

Step 4. *b*

From this assumption and Step 3, using modus ponens we derive:

Step 5. O*a*

Since any form of murder, gentle or otherwise, is still murder, then if Smith murders Jones gently, then Smith murders Jones:

Step 6. *a* ⊃ *b*

Then, by Steps 5, 6, and D2,

Step 6.5. O*a* ⊃ O*b*

Then, from Step 5, and modus ponens:

Step 7. O*b*

Finally, from Step 7 and applying D3 to Step 1, we see that there is a contradiction:

Step 8. (¬O¬*b*) . (O(¬*b*)).

Stated in plain English, this says, "it is and it is not obligatory that Smith murders Jones." This regimentation makes clear Joaquin's diagnosis of the argument: the paradox does not arise because there is something incoherent about deontic logic; rather, it is generated by a fundamental mistake in the formulation of its premises: as written, it treats the deontic operator O as ranging over indicatives of actions - statements describing what an agent does - when in fact, it can only properly range over imperatives of actions - statements commanding an agent to do something.

Joaquin emphasizes that the crux of the paradox lies in assuming that the obligation operator can be used on indicative statements, such as those of the form "S does A." Such propositions are statements of fact; they are descriptive and not normative. He argues that indicatives and obligations are two fundamentally different entities: facts can be true, false, necessary, etc., but they cannot be obligatory, while obligations cannot be treated as factual assertions. Thus, reading an expression such as

O (Smith murders Jones gently)

as "It is obligatory that Smith murders Jones gently" is already conceptually mistaken. It treats a descriptive proposition, a statement about something that happened, as an obligation, which is applying the O operator to the wrong type of statement.

He explains this by arguing as follows. The statement "Amanda cleans her room" describes the fact that Amanda cleans her room. However, if we are allowed to apply O to indicative statements, then

O (Amanda cleans her room)

is read as "It is obligatory that Amanda cleans her room." However, he objects that the fact that Amanda cleans her room cannot be obligatory. Intuitively, it does not make sense that facts can be obligated to happen. To avoid this

problem, Joaquin argues that the obligation operator cannot be applied to indicative statements, because doing so would be turning descriptive facts into normative prescriptions.

Joaquin proposes that the scope of the deontic operator should be limited to only imperative statements of the form "S, do A," or a statement of a mandate directed at an agent, rather than a description of what the agent does. For example, instead of the odd formulation

O (Amanda cleans her room),

the operator is properly applied to commands such as

O (Amanda, clean your room!).

He thinks that obligations should attach to acts an agent may or may not perform, not to states of affairs that may or may not happen, independently of the agent's actions.

How does identifying this problem provide Joaquin with a solution to the Gentle Murderer paradox? Given that he thinks the O operator attaches only to imperatives, he finds fault with the second principle of deontic logic - that $\boldsymbol{p} \supset \boldsymbol{q}$ entails $O\boldsymbol{p} \supset O\boldsymbol{q}$ doesn't make sense when applied to certain indicatives *p* and *q*. He gives the following example: "If Amanda's room is dirty, then Amanda cleans it" is a conditional statement in terms of indicatives, but it does not make sense to say that this statement entails "If it is obligatory that Amanda's room is dirty, then it is obligatory that Amanda cleans it." This is because a statement of fact such as "Amanda's room is dirty," according to his reasoning above, cannot be obligatory. This allows him to propose his "simplest solution" to the deepest paradox of deontic logic: if the paradox is stated in terms of indicatives, then it simply doesn't make sense, in the same way that "it is obligatory that Amanda's room is dirty" is ill-formulated.

Given the explanation above, Joaquin concludes that he has successfully defended his solution to the Gentle Murderer paradox. At the end of his paper, he briefly considers the following objection, which he calls a "riposte": What if the Gentle Murderer paradox were reformulated in terms of imperatives and not indicatives of actions? In particular, what if the controversial inferential rule (D2) were reformulated with its scope narrowed only to imperatives—would that not solve the problem?

In response, Joaquin briefly considers a revised version of D2, which he calls D2*:

(D2*) $(p! \supset q!) \Longrightarrow (Op! \supset Oq!)$,

where exclamation points following the sentence letters denote imperatives. He argues that using D2* in place of D2 blocks the inference from Steps 5 and 6 to Step 7, since we are no longer allowed to infer obligations from statements of fact, which D2 allows us to do in the original paradox. The introduction of D2* leads him to two cases:

- In the first case, granting Step 5 and D2* in the revised paradox, it is impossible to infer from the indicative contents of Step 6 the deontic statement in Step 7, since in the revised paradox, there is no deontic principle that allows for facts to entail obligations.

• In the second case, the inference from Step 5 to Step 7 in the paradox no longer works since D2* only applies to imperatives of actions, not indicatives.

Therefore, Joaquin concludes that writing a new rule, D2*, to be used only for imperatives, is sufficient to address the riposte, and that his original diagnosis about the problem with the paradox is correct. I don't find this response satisfactory, especially since Joaquin proceeds directly to rewriting the paradox using D2* instead of first taking the time to explain exactly what he means by a statement such as $p! \supset q!$. I also don't think that he can so easily dispense with the paradox by modifying just one of the deontic axioms - to construct a proper response to the objection, I think Joaquin should have revised all of them, making sure that O is applied only to imperatives. In the section below, I flesh out his argument, incorporating two kinds of sentences - imperative and indicative - to show why it fails.

EVALUATING JOAQUIN'S SOLUTION TO THE PARADOX

While Joaquin's use of formal methods to regiment the original statement of the Gentle Murderer Paradox is illuminating, his attempt to regiment the riposte and his short reply to it are inadequate at best. While he thinks that he can solve the entire paradox by revising the second axiom of deontic logic to represent imperatives and thus restrict the scope of application of O, he fails to consider that by doing so, he is implicitly using a system of deontic logic that combines indicative sentences with imperative sentences, which he entirely fails to discuss. If he were to flesh out the details of such a system, he would see that restricting the scope of the deontic operator to imperatives does not, in fact, solve the paradox. He would also see that while a statement consisting of entirely indicatives entailing a statement of entirely imperatives in D2 is a problem, an equally important problem is caused by the way the O operator is used on indicatives in D1, something he argues against in words but fails to formally regiment.

I think that in order to adequately address the objection raised in the riposte, he would need to describe the particular system of Deontic logic he intends to use in more detail, including how the classical logical operators should apply to imperatives and combinations of imperatives and indicatives, and how this would change the three Deontic axioms when applied to the paradox. While from his paper, it is not totally clear how his proposed system should work or what combinations should be allowed, I propose one plausible candidate for a system, and attempt to rewrite the paradox using this system.

My proposed hybrid logical system contains both indicatives, which are denoted by the usual variables p, q, r, etc., and imperatives, which are denoted by sentence letters in square brackets: $[p]$, $[q]$, $[r]$, etc. The connectives for propositional logic are defined in the usual way for indicative sentences, but there are some differences when applied to imperatives. First, I focus on combinations of two imperatives.

• The negation of an imperative, $\neg[p]$, is an imperative. For example, the

negation of "stop!" is "don't stop!"

- The conjunction of imperatives, [*p*].[*q*], is also an imperative. For example, the conjunction of "stop!" and "turn around!" is "stop and turn around!"
- Similarly, the disjunction of imperatives, [*p*] ∨ [*q*], is also an imperative. For example, the disjunction of "stop!" and "turn around!" is "stop or turn around!"
- The material conditional, when applied to two imperatives, [*p*] ⊃ [*q*], is also an imperative. The intuitive reasoning for this is that just like the material conditional, it is possible to meaningfully interpret the horseshoe between two imperatives regardless of whether they are substantively related, due to the equivalence of $p \supset q$ and $\neg p \vee q$ via the implication equivalence rule. For example, while the sentence "If [you] stop! Then turn around!" may sound a bit strange, it makes more sense when converted to the following: "Don't stop, or turn around!" Even if *p* and *q* aren't substantively related, it is possible to interpret the material conditional between them. For example, "If [you] stop! Then take out the trash!" can be converted into "Don't stop, or take out the trash!" which still makes sense.
- The biconditional operator, when placed between two imperatives, simply means that the commands are equivalent: for example, [*p*] ⇔ [*q*] would represent something like the equivalence between "stop!" and "don't move!"

The real difficulty when defining a system of Deontic logic that uses both indicatives and imperatives is making sense of combinations of both types of sentences. This raises various problems. First, I am not sure what to make of sentences such as *p*.[*q*] or $p \vee$[*q*], the English equivalents of which would be something like "the stoplight is red, and stop!" or "the stoplight is red, or stop!". The difficulty arises when attempting to negate such statements, in deciding whether the negation should apply to one or both of the constituent statements. To me, the following are equally unintuitive ways of negating a sentence such as "the stoplight is red, or stop!":

- "It is not the case that the stoplight is red, or stop!"
- "The stoplight is red, or don't stop!"
- "It is not the case that the stoplight is red, or don't stop!"

Similarly, I do not know how to treat statements such as [*p*] ⇔ *q* beyond considering them automatically false, since a command cannot be equivalent to a statement of fact. Joaquin also gives no indication as to how he would deal with these cases, and since they are not relevant to the three axioms of Deontic logic he uses in his paper, I leave these questions for further consideration.

For the purposes of my analysis, I focus on the material conditional between an indicative and an imperative, since this is present in the first axiom. There are two cases: [*p*] ⊃ *q* and *p* ⊃ [*q*]. To me, [*p*] ⊃ *q* does not make sense, since the imperative on its own can give no information on whether or not it was complied with, so it cannot imply any sort of indicative consequence. The truth value of this sentence would then only depend on *q*, which defeats the purpose of even combining *q* with an imperative antecedent.

The more interesting case is *p* ⊃ [*q*], which corresponds to English statements

such as "If the stoplight is red, then stop!" which do make sense intuitively. While it would be straightforward to allow the same material conditional as in propositional logic between p and $[q]$, this raises several problems. First, would the negation of such a sentence apply to p only, $[q]$ only, or both of them? For example, it is unclear which is the correct negation of the statement "If the stoplight is red, then stop!":

- "If it is not the case that the stoplight is red, then stop!"
- "If the stoplight is red, then don't stop!"
- "If it is not the case that the stoplight is red, then don't stop!"

To me, none of these statements is clearly more correct than the others. Trying to clear up this confusion by converting the conditional to a disjunction and using DeMorgan's Law leads to the following:

$\neg(p \supset [q]) \rightarrow \neg(\neg p \vee [q]) \rightarrow p.\neg[q]$,

which, interpreted in English, would read "The stoplight is red, and don't stop!" which runs into problems described previously.

For all of the reasons stated above, I think it is useful to define a version of $\supset$ that can connect an indicative antecedent with an imperative consequent to produce another imperative, though a conditional one. I denote this new operator by using the symbol "/". It has the same meaning and is used in the same way as $\supset$, except for one important difference - the negation of $p/[q]$ is defined as follows:

$\neg(p/[q]) \equiv (p/\neg[q])$.

Using the above example again, the negation of "If the stoplight is red, then stop!" is defined to be "If the stoplight is red, then don't stop!"

Now, I will address the question of whether this framework allows imperatives and their combinations with indicatives to be assigned truth values. While I do not think that it makes sense for standalone imperatives, such as "Wash the dishes!" or "Clean your room!" to be truth functional, I do think it is possible to work with them in a truth-functional context through defining operators that first convert them into indicatives. The first way to do this is to use the O operator, which, as Joaquin argues, accepts only imperatives and turns them into indicatives of the form "this imperative must be satisfied." For example,

O(Clean your room!) → [Someone] is obligated to clean their room could be true or false, depending on the state of the world.

However, I think that in order to represent the Gentle Murderer Paradox more accurately, another imperative-to-indicative operator is needed. Accordingly, I define the operator '$', which converts imperative statements to statements of fact, which can then be assigned a truth value based on whether or not the imperative was complied with. For example: the $ operator would convert the command

$[Wash the dishes!]

to the indicative statement "The dishes have been washed," which is truth-evaluable. While both of the examples above were given in impersonal form, it is possible to interpret both as applying to particular people, for example

O(Amanda, clean your room!) *turns into* Amanda is obligated to clean her room.

$(Amanda, wash the dishes!) *turns into* Amanda has washed the dishes.

Next, using the two newly defined operators, I attempt to revise the formulations of the Deontic axioms D1, D2, and D3 in terms of both indicative and imperative statements. Joaquin's original modification of the second axiom, which he calls D2*, reads as follows when rewritten using the new notation:

(D2*) $[p] \supset [q] \Rightarrow O[p] \supset O[q]$.

While Joaquin does not explain why this is the best way to rewrite D2, I agree that it is better than the original statement,

(D2) $(p \supset q) \Rightarrow (Op \supset Oq)$,

which doesn't make sense for two separate reasons: the fact that O is applied to indicatives, and the fact that indicative statements entail statements about obligations. While I think that Joaquin should have done more work in either defending or attacking this modified version of D2, I will accept Joaquin's D2* in my regimentation to show that it does not solve the problem in the way he thinks it does.

I now focus on revising D1, the original statement of which was

(D1) $O(p \supset q) \supset (p \supset Oq)$.

Under my new framework, this does not make sense since the O operator is applied to the indicatives $p \supset q$ and q. I propose the following modification:

(E1) $O(\$[p]/[q]) \supset (\$[p] \supset O[q])$.

Since $\$[p]/[q]$ is defined to be an imperative, while $\$[p]$ and $O[q]$ are indicatives, this modified version of E1 doesn't face the same problems as the original D1. Also, though it is not particularly relevant, D3 can be rewritten as follows:

(E3) $O([p]) \supset \neg O\neg([p])$.

Having modified the three principles of deontic logic to accommodate imperatives, I use these axioms to regiment the Gentle Murderer Paradox again, being careful with the scope of the operators that I apply. Let r represent the imperative sentence "Smith, kill Jones gently!" and s represent the imperative "Smith, kill Jones!" Recall that the original two assumptions were

(A1) All kinds of murder are forbidden.

(A2) In a case where murder is unavoidable, murdering someone violently is worse than murdering someone gently.

From (A1) and (A2), we have

Step 1. $O(\neg[s])$

Step 2. $O(\$[s]/[r])$

From Step 2, using E1:

Step 3. $\$[s] \supset O[r]$

By assumption:

Step 4. $\$[s]$

From this assumption and Step 3, using modus ponens, we derive the following:

Step 5. $O[r]$

Since any form of murder, gentle or otherwise, is still murder, the imperative "Smith, kill Jones gently!" entails the imperative "Smith, kill Jones!":

Step 6. $[r] \supset [s]$

Then, by D2*,

Step 6.5. O[*r*] ⊃ O[*s*]

From Steps 5, 6.5, and modus ponens:

Step 7. O[*s*]

Finally, from Step 7 and applying D3 to Step 1, we see that the contradiction returns!

Step 8. (¬O[¬*s*]).(O(¬[*s*])).

Thus, I have shown that Joaquin's proposed solution to restrict the scope of O to only imperatives and rewrite the principles of Deontic logic accordingly does not solve the Gentle Murderer Paradox. As it turns out, when stated in terms of imperatives and when the corresponding new rules are used, the paradox remains!

In conclusion, I think that Joaquin is correct that the paradox is properly formulated in terms of imperatives, and he takes a step in the right direction by rephrasing D2 in terms of imperatives. However, this only partially solves the problem, as D1 also contains indicative imperative entailment, and it is not clear why he fails to address this and only focuses on revising D2. Still, Joaquin's partial failure in addressing the objection in the riposte raises many interesting questions about how Deontic logics should properly define combinations of indicative and imperative statements, which are important to consider. My proposed method is far from the perfect solution to this question, though it provides a clear way to show that Joaquin fails to satisfactorily solve the Gentle Murderer Paradox, making his proposed "simplest solution" not so simple after all.

BIBLIOGRAPHY

Joaquin, Jeremiah Joven. "The Simplest Solution to the Deepest Paradox of Deontic Logic." *Organon* F 30, no. 4 (2023): 312–322.

Diana Nichvoloda is a junior at Harvard College originally from Boonton, New Jersey. She studies Philosophy and Economics with language citations in Ukrainian and Chinese. Her philosophical interests include social philosophy and metaphysics, logic, and political philosophy. Outside of academics, Diana works at the Harvard Ukrainian Research Institute and also enjoys running, skiing, and learning languages.

AUTHOR'S NOTE

The idea for this paper was originally suggested to me by my teaching fellow in a course called *Formal Methods in Philosophy*, who first introduced me to the Gentle Murderer Paradox and encouraged me to explore it more deeply. His guidance was invaluable throughout the writing process, especially in helping me think more carefully about how to move between philosophical claims expressed in words and their formal logical representations using modal and deontic logic. The most challenging part of working on this paper was

diagnosing what exactly goes wrong with the author's proposed solution to the logical paradox, which seemed very convincing at first. Joaquin's proposal to restrict the deontic operator to imperatives initially seemed promising, but I soon realized that simply modifying one of the principles of deontic logic was not sufficient, and that a restructuring of the entire logical system was needed to accommodate imperatives in all of the deontic principles. My attempt to do this raised more questions than answers, because not only did it show that the paradox is not as easily solved as Joaquin thought, but also exposed other difficulties about how deontic logic should handle mixed systems of indicatives and imperatives. Thus, this seemingly clear-cut logic paper became much more open-ended than I thought at first. It led me to consider interesting questions about the proper scope of logical operators, how exactly to formalize and define truth values in a mixed logical system that contains both indicative and imperative statements, and more broadly, the limits of formalization in ethical reasoning.

A Space Alien Case Against the Davidsonian Theory of Communication: What If Humans Are Not the Most Rational Beings in the World?

Xuning Alina Gao
Kenyon College

FOREWORD

Editors: Annelise Durham Gross, Kayla Lee, Richard Zhang

One of the most influential claims in philosophy of language comes from Donald Davidson, who argues that there is no such thing as a completely untranslatable language. In his perspective, anything we cannot translate simply wouldn't count as meaningful communication or even rational thought. This position requires an incredibly tight connection between translation, language, and rationality. But what if that connection is too narrow?

The following paper challenges Davidson, and asserts that untranslatability doesn't have to work both ways. Central to the Gao's argument is Jakob von Uexküll's concept of umwelt, the idea that different beings exist in vastly different perceptual and cognitive "worlds." Armed with this idea, the paper reframes language as tied to a specific form of life rather than just a shared system of meanings. We were intrigued by the Gao's use of a creative edge case involving space aliens to test Davidson's view. Readers are prompted to ask themselves the following question: If we encountered an alien species, would we still recognize their intelligence and rationality even if we had zero way of translating their language? Gao suggests we likely would, based on how they behave and interact, even if their actual speech stays completely outside our reach.

As we live in a world shaped by constant scientific advancements and technological change, such an argument urges us to be more cautious about how we attribute a lack of intelligence or rationality. It applies just as easily to everyday cross-cultural communication as it does to emerging, cutting edge forms of artificial intelligence or the possibility of discovering new—and even alien—forms of life. At the same time, the paper reminds us that there may be forms of intelligence and rationality that we may not even begin to comprehend. In recognizing this, we hope that readers come away from this paper with a clearer sense of the limits of human understanding, and perhaps a little humbled.

ABSTRACT

Donald Davidson equates alternative conceptual schemes with untranslatable languages, arguing that a completely untranslatable language is an incoherent notion: if a language is completely untranslatable, its speakers cannot be recognized as rational agents or language users. This paper challenges that argument. I propose that both rationality and language can be attributed in cases of complete untranslatability by appealing to perceptual cues and complex, patterned behavior, which are grounds independent of translatability. This claim is supported by an argument from asymmetry: complete untranslatability need not be symmetrical between languages, and in asymmetrical cases, untranslatable languages can be recognized as such.

This paper is not a realist defense of conceptual schemes. It does not deny

that the recognition of an alternative conceptual scheme requires a common ground. Rather, the disagreement concerns what the common ground is: Whether linguistic data exhaust the attribution of rationality, language, and conceptual scheme. Davidson says yes, and I argue no. By broadening the evidential basis for recognition beyond semantic data, I hope to open up new ways of developing a theory of communication that accounts for aspects of our communicative practices that are left unaddressed by semantics, syntax, and pragmatics.

INTRODUCTION

Donald Davidson, on the basis of his papers "On the Very Idea of a Conceptual Scheme" (1973) and "Rational Animals" (1982), is committed to the view that beings whose language is completely untranslatable cannot be recognized as rational or as possessing a language. While he does not state this thesis explicitly, it follows from the conjunction of his arguments against complete untranslatability and his account of rationality. In the 1973 paper, he rejects the very idea of incommensurable conceptual schemes. Equating incommensurable conceptual schemes with untranslatable languages, he claims that a being with a completely untranslatable language cannot be considered as having a language at all. In the 1982 paper, he further argues that linguistic communication, and thus a translatable language, is the criterion of rationality attribution.

Taken together, Davidson argues that the attribution of both language and rationality requires a translatable language. I argue that this is mistaken: We can attribute rationality and language even in cases of untranslatability. In this paper, I defend the coherence of completely untranslatable languages. My defense for complete untranslatability proceeds in three steps: (a) Rationality attribution does not require a translatable language, and (b) in spite of Davidson's claims, complete untranslatability may not preclude successful communication. Eventually, I propose that (c) in cases of asymmetry, there are completely untranslatable languages that are recognizable as languages. If my argument is reasonable, Davidson's argument against conceptual schemes cannot be sustained, and we must adopt broader criteria of what counts as rationality and language.

Yet, this paper is not a realist defense of untranslatable languages. It does not deny the Davidsonian thesis that a common ground between languages is required for one to recognize an untranslatable language. Rather, the disagreement between me and Davidson concerns what the common ground is: whether semantic evidence exhausts the attribution of rationality, language, and conceptual schemes. Davidson says yes, and I argue no.

In Sections I and II, I explain Davidson's critique of complete untranslatability in relation to his argument on rationality attribution. In Sections III and IV, I argue against Davidson's criterion for rationality attribution, that even if *being* rational entails that one has a language, as Davidson argues, the *attribution* of

rationality to someone else does not entail that we can translate their language. It is possible to attribute rationality to another being without being able to understand their language. In Section V, I argue that untranslatability might be asymmetrical between languages, which allows communication despite complete untranslatability. Finally, such asymmetrical translatability challenges Davidson's criterion of languagehood.

I. DAVIDSON ON COMPLETE UNTRANSLATABILITY

A significant issue at hand is that the term "language" seems to be used in various ways throughout Davidson's 1973 paper.[169] Davidson then has two similar, but different, attacks towards the existence of an untranslatable language.[170] On the one hand, emphasizing "language" as a speech behavior, Davidson argues that "a form of activity that cannot be interpreted as language in our language is not speech behavior" (1973, 7). In other words, if one's expression is completely untranslatable to us, it is unrecognizable as an expression, and we will not regard one as speaking a language, but rather as making noise. Moreover, without being able to translate one's language, Davidson argues that we could not identify the speaker of the untranslatable language as behaving intentionally or as possessing beliefs, as he puts it: "It seems unlikely that we can intelligibly attribute attitudes as complex as these to a speaker unless we can translate his words into ours" (1973, 8).[171]

On the other hand, viewing "language" as an abstract collection of concepts, Davidson questions whether we could recognize an alien set of concepts as "largely true but not translatable" (1973, 16). Following Tarski's theory of truth, he argues that the totality of T-sentences—sentences considered to be trivially true by English speakers, such as "'snow is white' is true if and only if snow is white"—"uniquely determines the extension of the concept of truth for English" (1973, 16). Truth is merely "our best *intuition* as to how the concept of truth is used," or as Tarski calls it, the Convention T (1973, 16). A sentence is "true" only insofar as it is translatable to one of our T-sentences.

Such a deflationary account of truth makes it impossible to "divorce the notion of truth from that of translation" (Davidson 1973, 16). Conceding the impossibility for one to assess an alternative conceptual scheme while "temporarily shedding his own," Davidson thinks that even if an untranslatable alien language exists, English speakers will not recognize it as meaningful and patterned (Davidson 1973, 7). In other words, any sort of interpretation would put the "language" in question under the category of partial rather than complete untranslatability, as such interpretation necessarily involves what Davidson calls "the principle of charity"—to even start the process of translation, the

---

169 His concept of "rationality" also lacks a positive definition, which I will address later in this paper.

170 Although the paper is a critique on conceptual schemes, I wish to only focus on his argument against untranslatable languages and show why it is an intelligible notion.

171 "Unlikely" in the quote sounds like a statistical claim, but I consider Davidson to be making a nonstatistical claim.

interpreters need to attribute "conditions of truth that actually obtain (in our own opinion) just when the speaker holds those sentences true" (1973, 19). Yet, once it is under the category of partial untranslatability, it is no longer the subject of discussion for Davidson's critique towards complete untranslatability. Because the alien language in question is completely untranslatable, Davidson argues that the English speakers will struggle to find the meaning of the aliens' sentences as coherent and comprehensible, and there is no reason to regard it as a language at all. Davidson then claims that a language with a completely different set of T-sentences is *epistemologically unrecognizable* by us, which renders it *metaphysically impossible*.

The two lines of attack on complete untranslatability come down to the same thesis, that: (i) The criterion of languagehood is 'translatability into our language.' If an action or a set of concepts is not translatable to our language at all, we will not recognize it as a language. From now on, I will refer to this thesis as Davidson's criterion of languagehood. In the following, I will object to Davidson's criterion of languagehood by focusing on Davidson's former attack concerning 'language as a speech behavior.' The critique, I hope, would eventually hold his latter attack as positing an *incomplete* account of complete untranslatability—Davidson considers distinct sets of T-sentences as the only source of complete untranslatability. I will argue otherwise.

II. COMPLETE UNTRANSLATABILITY IN CONVERSATION WITH RATIONALITY

Davidson's first critique regarding language as speech behavior already faces objections. For instance, P. M. S. Hacker argues that it is possible for one to identify a language-using community without translating their language—we can refer "to their way of life or even just to the artefacts they produce" (1996, 291). Hacker raises the Etruscan civilization as an example, that although we have not been able to fully translate Etruscan, "there is no doubt that Etruscans spoke a language and that the surviving inscriptions are indeed such" (1996, 291). In other words, Hacker argues against Davidson's criterion of languagehood that, from one's complex, patterned behavior, it is possible to infer and recognize one's possession and usage of a language independent of translating the language. As Hacker claims, "we should identify the forms of behaviour and behavioural capacities that constitute criteria for speaking a language" (1996, 307). Davidson's argument that an untranslatable language cannot be identified as speech behavior then appears to fall short.

Davidson may have a response to Hacker in his 1982 paper "Rational Animals," which draws upon the connection between language, complex behavior, and rationality. Davidson argues that "there is a [complex] pattern [of behavior] only if the agent has a language" (1982, 322). In his view, genuinely complex behavior—the behavior to which we attribute propositional attitudes—presupposes a translatable language, as to have propositional attitudes (such as beliefs and desires) means to possess concepts, which entails the possession of

a (translatable) language.[172] Moreover, such complex behavior, as it entails one's possession of a translatable language, indicates one's rationality, that: "unless there is actually such a complex pattern of behavior, there is no thought" (Davidson 1982, 322).

Davidson does concede that not all behaviors we recognize as patterned indicate that one possesses propositional attitudes and thus a translatable language. It is possible for a creature to "react with the world in complex ways without entertaining any propositions," such as discerning colors, increasing its food intake, or preserving its life (1982, 326). However, Davidson claims that none of such behaviors "shows that the creature commands the subject-object contrast," which demonstrates the creature's capability of possessing beliefs and thoughts (1982, 326).[173] Rather, linguistic communication, and with it, the possession of a translatable language, is a necessary condition for describing one's behavior as being genuinely complex. Without identifying a translatable language, our attribution of a being's behavior as sufficiently complex would either "be wrong" (Davidson 1982, 324) or would be applying an explanation "far stronger than the observed behavior requires" (Davidson 1982, 325). The attribution would be wrong in the same way as attributing thoughts and attitudes to a heat-seeking missile is wrong; the attribution would be stronger than required as in the case of attributing beliefs to speechless animals who are incapable of thought. Davidson's claim is thus: (ii) If one can be recognized as having complex behavior, then one must have an in principle translatable language (but not one that we can necessarily translate).[174]

Davidson could then respond to Hacker with "Rational Animals": Yes, we can recognize the community's behavior as complex and patterned, and attribute propositional attitudes to the community members. However, such recognition and attribution are not the preconditions, but the consequences, of the community having a translatable language. To use Hacker's example, the fact that we can recognize the Etruscan community as having a complex patterned way of living indicates that Etruscan must have a translatable language. Our failure of translating Etruscan language does not arise from completely non-overlapping concepts between English and Etruscan, but arises from the lack of means and materials. Had there been a community possessing a completely untranslatable language (a non sequitur for Davidson), we would not have succeeded in viewing its behavior as complex or attributing propositional attitudes to it. In other words, a case of successfully translated speech behavior confirms the translatability between languages, while an instance of translation

---

172 As a matter of fact, for Davidson, there is no such thing as an untranslatable language, but I will continue using the term "translatable language" for clarity.

173 It appears that research on animal behavior is showing it to be more complex than Davidson thought in 1982. Whether our current understanding of animal behavior shows that it is sufficiently complex to Davidson's standard requires another paper.

174 Davidson seems to be committed to the claim that Etruscan language, though not translated, is translatable. It is not translated because we lack sufficient materials and resources, but is still an in principle translatable language.

failure does not necessarily lead to complete untranslatability.[175] A community's behavior is regarded as intelligible and complex only if it has a language that is translatable to ours.

In conceding this point, Davidson appears to acknowledge an alternative to the criterion of languagehood, that one's genuinely complex behavior entails one's possession of a language. Yet, at least in the 1982 paper, Davidson does not provide clear arguments for what counts as complex behavior except that it entails a translatable language. I do not wish to delve into this issue, but I do challenge Davidson's claim that the recognition of complexity requires recognition of a translatable language. Eventually, such an objection enables me to argue in agreement with Hacker, that complex behavior does not guarantee a translatable language, and linguistic translatability is not the only way to recognize the community as possessing a language.

Such an argument is deeply interwoven with Davidson's conception of rationality. In both articles (1973, 1982), Davidson claims that rationality and the possession of a translatable language is inseparable, that "language is not a trait a man can lose while retaining the power of thought" (1973, 7), and "rationality is a social trait. Only communicators have it" (1982, 327). In simple terms, Davidson thinks to have rationality is to have thoughts and beliefs, and to have a belief is to be capable of distinguishing between subjects and objects. The possession of such distinction is only possible through having a language. As a deflationist on rationality, Davidson argues that one's rationality can only be attributed through linguistic communication, which depends upon both communicants having "the concept of an intersubjective (objective) world [...] about which each communicant can have beliefs" (Davidson 1982, 327). Davidson then argues for: (iii) If one *possesses* rationality, then one must possess a language, and (iv) One can be *recognized* as rational if and only if one has a translatable language. I will refer to (iv) as Davidson's criterion of rationality attribution throughout the rest of this paper.

"Rational Animals" thus strengthens and clarifies Davidson's attack on untranslatable speech behavior with regards to both complex patterned behavior and rationality: If one's language is completely untranslatable, we would not be able to recognize any of one's behavior as complex, or as having intentions or beliefs—their action would appear to us as crazy and confusing; we also would not recognize one as rational due to one's lack of a translatable language and incapability of linguistic communication. *This criterion of rationality attribution is what I am attacking.* I will argue that there could exist beings whose behaviors we unmistakably perceive as intentional and purposive, which lead us to view them as possessing complex behavior and rationality. Yet, we find no evidence that they possess a translatable language. For the sake of argument, I grant Davidson's definition of rationality, that (iii) if one possesses

175 The apparent circularity of this position comes from the fact that Davidson treats these concepts as interrelated and interwoven.

rationality, then one must possess a language.[176] But I challenge Davidson's criterion of rationality attribution—a being to whom we *attribute* rationality does not necessarily possess a language that is potentially *translatable* to ours.

III. WHAT IS AN *UMWELT*?

My argument that rationality does not require translatability relies on the concept of *umwelt*. Developed by Jacob von Uexküll and popularized in his paper "A Stroll Through the Worlds of Animals and Men," an *umwelt* is a closed unit formed by the "perceptual and effector worlds" (1934, 6). For humans, the perceptual world includes both our senses and the tools that extend them, such as telescopes and spectacles. The effector world, on the other hand, is "all that [a subject] does" (Uexküll 1934, 6). The effector tools involve not only our limbs, but also transportation, such as airplanes, and tools, such as knives and guns.

Every species has an *umwelt*. Against the behaviorists' and mechanists' canon, Uexküll argues that just like we grant humans as active agents, as "operators," with our own perspectives of the world, animals and insects should not only be merely seen as "a collection of perceptual and effector tools," but considered as having operators underlying those perceptions and actions (1934, 6). Animals are experiencing their own perceptual-effector *umwelten*, which "is only a section carved out of the environment which we see spread around it," and the environment itself is nothing but part of our human *umwelt* (Uexküll 1934, 13). As different subjects have different receptor cues, perceptions, sensory of time, search tones, familiar objects, and more, their *umwelten* vary in space, time, and qualities of the objects within them. To give an example, as snails have relatively slow receptor time, the indivisible moment that they can register is much longer than the indivisible moment of human beings. Thus, a rod that appears to humans to oscillate four times per second is stationary in a snail's *umwelt*.

In a word, due to beings' different physical constructions, they see different things and exist in different *umwelten*. As Uexküll argues, "all animals, from the simplest to the most complex, are fitted into their unique worlds with equal completeness. A simple world corresponds to a simple animal, a well-articulated world to a complex one" (1934, 11).

IV. RECOGNIZING RATIONALITY WITHOUT TRANSLATABILITY

To quickly recapitulate my argument, Davidson assumes that if we are able to recognize a being's behavior as complex, it must be that they have a translatable language. In such a case, we attribute rationality to them. Yet, I argue that the criterion of rationality attribution is not translatability. It is important to note

---

176 I do not wish to participate in the debate of whether animals who demonstrate highly intelligent puzzle solving skills and behaviors, such as bees and octopuses, are rational. I do not dispute Davidson's claim that rationality requires possession of a language, what I am arguing against is his definition of how we recognize one as rational.

that I am here addressing rationality attribution, which Davidson claims requires actual translatability, not in principle, potential translatability.[177] He holds that potential translatability is required by recognizing complex behavior, which I will address later in this paper.

My central claim is that it is possible for there to exist beings whose *umwelt* is structurally more complicated than ours. This is possible in many forms, but I use a science-fiction example for illustration. Let us suppose the beings are multi-dimensional creatures. Let us also suppose that they only use a language that exists in a dimension outside of the human *umwelt* to communicate with each other. Thus, we humans are able to perceive their physical, three-dimensional existence and physical interactions with each other, while our perceptual tools are utterly insufficient to access their language. Yet, insofar as humans can observe, these beings cooperate with each other, have well-formed artefacts, and operate sophisticated technologies.

An example might help clarify the possibility of such beings' existence. Imagine a group of humanoid space aliens arriving on Earth in a spaceship. The aliens are of an observable size, and we observe them moving off of the spaceship, pressing buttons, and manipulating complex-looking machines. Oddly enough, although it is apparent that the aliens demonstrate cooperative behavior, there is no sign of them performing speech behavior—no sounds are heard and no signs are seen, not even eye contacts are made (if they even have eyes). Suppose humans attempt to interact with the aliens, there could be different possibilities of how things go. Through years of studying, humans may be able to translate the aliens' language. This would be the case when their *umwelt* is penetrable through human perceptual tools.

Yet, I would say there is another possibility, that the aliens have a much more complicated *umwelt*, and their way of communicating with each other falls outside of human *umwelt* altogether, beyond our perceptual and conceptual grasp. Suppose the aliens are six-dimensional creatures and their language only exists in the fifth dimension, humans could not perceive the aliens using language, even though their speech behavior happens right next to humans in space-time. As established above, Davidson concedes (iv): One can be recognized as rational if and only if one has a translatable language. Given that no matter how hard we try, we are unable to perceive any sign of the aliens having a language, such a criterion of rationality attribution leads us to conclude that because we fail to identify a translatable language, space aliens are not rational.

Davidson's conclusion is controversial. We can clearly observe that the aliens are able to manipulate highly sophisticated spaceships, travel to places that humans cannot, and construct far more advanced technologies than humans can. Such cases indicate that the aliens are highly complex and intelligent. While there is a reasonable debate about how to measure the complexity of

177 Actual translatability indicates a language that is translated, while potential translatability means a language can be translated in principle, but for various reasons, such as the lack of materials or the non-cooperation of the speaker, it is not translated.

creatures' behavior, if human behavior qualifies as complex for Davidson, then the aliens' behavior should too. Admittedly, the boundary of complexity would remain fuzzy in borderline cases. Yet, when contrasting the actions of aliens operating spaceships with those of rabbits hopping about eating hay, the distinction seems uncontroversial. We may then attribute rationality to the aliens from their sufficiently complex behavior alone. Thus, Davidson's criterion of rationality attribution would deny the rationality of a being who we would clearly consider rational. Consequently, his criterion of rationality attribution—one can be recognized as rational if and only if one possesses a translatable language—is unsatisfactory, and translatability of languages is not necessary for rationality attribution. Moreover, without providing an account of what rationality amounts to, Davidson risks conflating rationality and translatability by giving a circular and rather trivial argument. He seems to only be saying "Having a translatable language (translatability) is the criterion for having a translatable language (rationality)." Separating genuinely complex behavior from translatability and treating it as a stand-alone criterion for rationality provides the notion with substantive content, that it does not merely mean possession of a translatable language, but also stands for one's capability of being intelligent, solving problems, responding to reasons, and more.

Davidsonian critics might object through Davidson's (ii) claim, that if one can be recognized as having complex behavior, then one must have a potentially translatable language. Once we are seeing the aliens as rational and their behavior as complex, interpretation has taken place and the principle of charity is necessarily involved. The alien example given above is then operating under the category of partial, rather than complete, untranslatability. Nonetheless, viewing the aliens as rational does not mean their language is partially translatable. There are three ways of answering this objection. First, such an objection arises from the Davidsonian circular presupposition that rationality attribution requires translatability. We are now at an impasse—the objection and my response are results of a more substantial difference in our criterion of rationality attribution. Yet, as established above, compared to Davidson's translatability criterion, my criterion—complex behavior alone is sufficient for rationality attribution—provides a more satisfactory explanation for our intuitions regarding the rationality of speechless intelligence. It also provides a more substantive account of rationality that extends beyond one's possession of linguistic tools. Thus, the presupposition that rationality attribution requires translatability is not justified, and the aliens are not under the category of partial untranslatability.

My second response to this objection is that the critics are making a category mistake by conflating "interpretation of behavior" with "translation of language." While we may assume their actions are goal-directed and highly intelligent, our translation of their linguistic symbols in the fifth dimension is by no means guaranteed. The principle of charity is a theory of meaning that operates within the linguistic realm, and the interpretation of behavior, though it could be charitable, does not involve the (semantic) principle of charity itself.

Therefore, the aliens' rationality is not a byproduct of their language being partially translatable, but an inherent quality of their behavioral complexity.

My last response addresses the definition of untranslatability. For a language to be "in principle translatable" simply means that the language-users have sufficiently similar sets of

T-sentences that, given enough time and effort, allow them to interpret and translate each other's utterances. Yet, because the alien language exceeds the human *umwelt*, any of the aliens' speech behavior would be beyond human perception. To translate the alien language would then require us to construct a theory of meaning (interpret their Convention T) without any perceivable empirical input. Moreover, with a much more complicated *umwel*t, it is unlikely that the aliens' T-sentences would be similar to ours in any way. Thus, the aliens' behavior, such as operating a spaceship, can be recognized as sufficiently complex without implying their possession of an in principle translatable language. The example is not a case of partial untranslatability, and genuinely complex behavior is a sufficient criterion for rationality attribution.

V. LANGUAGEHOOD WITH ASYMMETRY

In the section above, we discussed Davidson's criterion of rationality attribution, and now we move on to discuss his conception of languagehood. For Davidson, rationality requires a translatable language, and language, for Davidson, is always translatable. It is the second claim—there is no untranslatable language—I am critiquing here. In his 1973 paper, Davidson argues: (i) The criterion of languagehood is 'translatability into our language.' With the case of the space alien, Davidson would then say: As we fail to recognize any trace of the aliens' language, why consider them as having a language at all? I challenge Davidson's criterion of languagehood. I will argue that we can recognize the aliens' possession of a language though it is completely untranslatable. For Davidson, a completely untranslatable language being counted as a language is incoherent. I will present two arguments against Davidson's claim.

A. Identification through Recognition of Rationality

First, in the space alien example, I have argued that the aliens' highly complex behavior and interactions can serve as sufficient evidence for us to recognize their rationality, even though we fail to translate their language. I claim to have established that they are rational. I have also claimed to grant Davidson's conception of rationality, that: (iii) If one *possesses* rationality one must *possess* a language, then we should conclude that they must *possess* a language, although we utterly fail to translate it. Thus, it follows that we can recognize one's possession of language from one's rationality alone, independent of the language's translatability.

B. Identification through Testimony

The above argument against Davidson's criterion of languagehood might appear to be too swift, or that some may still doubt the aliens' possession of a language, though granting their rationality. I now provide an additional account. As Davidson thinks one cannot assess an alternative conceptual scheme while "temporarily shedding his own," he fails to consider the cases of asymmetrical complete untranslatability: If complete untranslatability were to exist, would it necessarily apply to both languages (the translating language and translated language) involved (1973, 7)? I think the answer is no. There can plausibly be cases of asymmetrical untranslatability. I now provide three cases of complete untranslatability, with the first case being the complete untranslatability that Davidson has in mind. The latter two cases, which Davidson fails to address, demonstrate how complete untranslatability could be asymmetrical. I argue that in the third case, we can recognize the existence of a completely untranslatable language.

Case 1: Davidsonian Complete Untranslatability

The first possible case of complete untranslatability would be when we English speakers encounter some beings who are making sounds and we cannot translate their sounds or see them as meaningful whatsoever. If we were Davidsonians, we would not consider these beings as having a language at all and would not consider this as a scenario where we encounter a case of complete untranslatability. In such a scenario, we have a complete breakdown of communication, or as Davidsonians would put it, there is no one to communicate with in the first place.

Case 2: Diachronic Complete Untranslatability

The second case of asymmetrical complete untranslatability would be when we encounter inscriptions and texts of a historical community. The English speakers are able to translate the community's language, while the community utterly fails to translate ours. The asymmetry comes from a lack of epistemic opportunity due to differences in spatial-temporal positions. A real-life example would be the Vikings. With the current intellectual development and linguistic knowledge, English speakers are able to translate the Viking language by analyzing their inscriptions; while the Vikings, without any access to modern English, would find English to be completely untranslatable—English did not exist at that time period. Davidsonians would say (uncontroversially) that we can recognize the Vikings as rational beings with a language, while the Vikings could not even recognize our existence, not to mention our rationality or possession of language. Yet, the Vikings' complete untranslatability of English does not undermine the English speakers' rationality and possession of a language—had the Vikings encountered the English speakers, it is not

controversial to argue that they, too, would view us as language- and rationality-possessing beings. Although this second case of untranslatability may not be a case of communication, it shows that one party's complete untranslatability does not necessarily imply that the other party lacks a language.

Case 3: Umwelten Complete Untranslatability

Building off of the second case, I now move on to the third and most important case of asymmetrical complete untranslatability: It is possible for English speakers to find the other beings' behavior to be completely untranslatable, while the other beings are able to translate English. Such a case can take many forms, but I will discuss the asymmetry that arises from *umwelten* differences. To revisit the space alien example, since the *umwelt* of English speakers is a part of the aliens' *umwelt* (three of the six dimensions), to some extent, we share the same (parts of the) world. The aliens have the perceptual tools that English speakers have, and can thus "share a concept of truth" with the English speakers (Davidson 1982, 327). It is then possible for the space aliens to recognize English speakers as having a language and learn English through observing English speakers' responses to three-dimensional stimuli, and communicate with English speakers in English.[178] Yet, the aliens' language would remain completely untranslatable to English speakers. Even if the aliens attempt to teach us their language, the language would have, what we English speakers call "meanings," that are rooted in their *umwelt* but not ours, which renders their language untranslatable to English. However, as the aliens can communicate with us in English, complete untranslatability does not necessarily entail a complete breakdown of communication.

It is plausible that the aliens with a more complicated *umwelt* could translate and learn English. Humans are able to have a good understanding of the ways of communication or signaling among beings whom we consider to have lower intelligence, such as bees' waggling dance. Similarly, the aliens are able to interpret humans' way of interacting with each other through observing our reaction to the environment and responses to each other. Another comparable everyday example is the communication between a three-year-old child and an adult. The adult can understand the child and interpret the child's utterances as translatable. The adult can communicate with the child in simple terms that the child can comprehend. Yet, due to the child's limited conceptual capacity and the underdeveloped brain complexity, the three-year-old child is unable to translate the adult's language. A similar case of child-adult asymmetry has been raised by Thomas Nagel (1986) when defending his realism. Although the child-adult case might arguably not be a case of complete untranslatability, the same reasoning applies to the case of English and the aliens' language (call it X)—even though X remains completely untranslatable from the perspective

178 It is questionable whether the aliens with six-dimensional existence would even consider English as a language, or merely as a way to signal each other. In the same way as we consider animal communication as mere signaling. Either way, communication is guaranteed.

of us English speakers, communication between the English speakers and the aliens is preserved in English. Yet, just as the Vikings' inability to translate English does not undermine the English speakers' possession of a language, our complete untranslatability of X does not imply that the aliens lack a language of their own beyond English. Together, the second and third cases argue against the Davidsonian criterion of languagehood as applied in the first scenario, that if one's language is not translatable to us, they do not have a language. My point here has a parallel with Ludwig Wittgenstein, that I treat language not as a stand-alone system, but as part of a life, which I refer to as *umwelt*.

The aliens' testimony then provides another challenge to Davidson's criterion of languagehood. In a case where the aliens are able to communicate with the humans in English, it is possible for the aliens to tell us in English that they have a language that lies outside humans' *umwelt* and is thus ungraspable by us. In such cases, given the aliens' highly intelligent and complex behavior and our past success in communication, it is unlikely that the humans persist in doubting the aliens' possession of a language, even though it is untranslatable. Through trusting the aliens' English testimony, we can recognize the beings with a completely untranslatable language as having a language.

VI. POTENTIAL DAVIDSONIAN REBUTTAL AND CONCLUSION

Some may argue that as the aliens are able to communicate with us in English, Davidson would agree with my claim that the aliens are rational and possess a language. Even if such agreement is granted, it does not undermine my argument. My (a) argument, that rationality can be attributed without linguistic communication, is separable from my (c) argument, that languagehood can be attributed without translatability. Consider, then, a case in which the aliens relevant to (a) never acquire or use any Earth language to communicate with humans. Given their sufficiently complex behavior, we can still identify them as rational even in the absence of linguistic communication. The attribution of rationality therefore does not depend on translatability or shared language.

Moreover, even in cases where aliens do speak English, I doubt Davidson would agree with me that English speakers can recognize the aliens as possessing a completely untranslatable language. According to Davidson, such recognition would require translatability, and by hypothesis, English speakers fail to translate the aliens' propositions. Thus, while Davidson and I may agree that the aliens speak English, we disagree on whether one's possession of a completely untranslatable language can be recognized through testimony alone.

Although Davidson does not participate in any direct discussion on the testimony case, he addresses an adjacent scenario in his 1973 paper on conceptual schemes. The scenario is built upon the idea that the "enough translatable difference may add up to an untranslatable one" (Davidson 1973, 8). Davidson suggests that there might be two languages, Saturnian and Plutonian. Saturnian is conceptually closer and translatable to English, while Plutonian is further away and completely untranslatable to English. Yet, suppose Plutonian

is translatable to Saturnian, and the Saturnian speaker tells us that they are now translating Plutonian. Davidson proposes that in this case, Plutonian is so alien to us that we would question "whether our translations of Saturnian were correct" (1973, 8). It is debatable whether we would respond that way.

Besides the fact that I find Davidson's response to this scenario unconvincing, we can infer that he might respond to my asymmetrical one-directional communication case that if the aliens tell us they have a language completely beyond our reach in English, we would start to question whether the aliens are speaking proper English. Nonetheless, this account is highly implausible. Having observed the aliens' highly rational behaviors, such as manipulating spaceships, and having communicated with them multiple times in English, it would be hard not to believe that the aliens speak proper English. With the space aliens' appearance (i.e. being green) and advanced technology, their claim that they exist in higher dimensions and possess languages untranslatable to us seems convictive. In other words, we regard certain beings as rational and as possessing a completely untranslatable language only if they fit our stereotypical vision of intelligent alien beings.

In conclusion, this paper argues against Davidson's interwoven view on rationality, complex behavior, and translatability. What the Davidsonian framework is missing is the role of *umwelt* in rationality attribution and language recognition. Through emphasizing the role of the *umwelt*, we can recognize one as rational without finding their language translatable, and we can recognize one as having a language without their language being translatable. As a result, Davidson's criterion of languagehood is overly restrictive, and his rejection of conceptual relativism is unsustainable.

BIBLIOGRAPHY

Davidson, Donald. "On the Very Idea of a Conceptual Scheme." Proceedings and Addresses of the American Philosophical Association, vol. 47, 1973, pp. 5–20. JSTOR, https://doi.org/10.2307/3129898. Accessed 19 Nov. 2025.

Davidson, Donald. "Rational Animals." *Dialectica*, vol. 36, no. 4, 1982, pp. 317–27. JSTOR, http://www.jstor.org/stable/42969807. Accessed 19 Nov. 2025.

Hacker, P. M. S. "On Davidson's Idea of a Conceptual Scheme." *The Philosophical Quarterly*, vol. 46, no. 184, July 1996, pp. 289–307. EBSCOhost, https://www.jstor.org/stable/2956442. Accessed 30 Nov. 2025.

Nagel, Thomas. *The View from Nowhere*. Oxford University Press, 1986.

Uexküll, Jakob von. "A Stroll through the Worlds of Animals and Men." German Essays on Science in the 20th Century, pp. 171–78. EBSCOhost, research.ebsco.com/linkprocessor/plink?id=8d073e4e-f1bc-35d4-9715-4ba56972b983. Accessed 13 June 2025.

Xuning Gao (also known as Alina) is a senior philosophy major at Kenyon College in Ohio (class of 2026). She spent a semester abroad at Pembroke College, University of Cambridge, studying philosophy of mind and metaphysics. Her current research interests lie in (social) epistemology and philosophy of language. Specifically, she explores how individuals (both human and non-human agents) interpret one another, how our evolving technological means shape our conception and perception of the world, and what factors are involved in communication. Outside of philosophy (if there is an outside), Xuning enjoys cooking, playing various sports, mingling with classical music, and watching weird movies.

AUTHOR'S NOTE

This paper aims to achieve two goals. First, I wish to defend the coherence of the idea of a conceptual scheme and Kuhnian conceptual relativism under an anti-realist picture. Second, I hope to challenge the Davidsonian language-centered framework and propose that our communicative practices cannot be fully understood without accounting for physical, perceptual, and social factors. This paper of mine is adapted from my senior Honors Thesis, titled "Communication with the Incommensurable: A Case Against Davidsonian Semantics and for Untranslatable Languages," through which I paint a fuller picture of my overall research project.

The inspiration for writing this paper has both an academic and a personal origin. Academically, I have always been interested in the theory of communication and interpretation. In my undergraduate years, I studied under two brilliant philosophers, Dr. Alexandra Bradner and Dr. Joel Richeimer (both of whom I secretly call my "academic family"). They introduced me to the conceptual scheme debate from different angles (i.e. Donald Davidson and W.V.O. Quine vs. Thomas Kuhn and N.R. Hanson). My intellectual conversations with them have always been vibrant, exciting, and memorable, and I am deeply grateful for their guidance along my path into academia.

On a personal level, as an international student from China, I face numerous obstacles in expressing and interpreting ideas, which, to me, indicates systematic differences between the two languages, or modes of thinking. My international background led me to deeply sympathize with the concepts of "incommensurability" and "conceptual schemes"—terminologies that cleanly accommodate my everyday experiences. Believing they are useful concepts that capture something fundamentally true about our communicative practices, I wish to defend them against Davidson's attack.