

# Logos : λόγος

Volume XVII • Spring 2021

Copyright © 2021  
Logos: The Undergraduate Journal of Philosophy  
logos.philosophy.cornell.edu  
Cornell University Ithaca, NY

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

# Logos : λόγος

Volume XVII • Spring 2021

*Editor-In-Chief:*

Bertrand Li

*Club President:*

Weston Barker

*Vice President:*

Liam Gale

*Treasurer:*

Rebecca Sabin

*Secretary:*

Fatima Yousuf

*General Editorial Staff:*

Jacqueline Allen

Fabio Cabrera

Ashley Gasdow (*Emeritus Editor-In-Chief*)

Meghana Gavirneni

Edward Gu

Zoey Hu

Ethan Kovnat

Ben Mendenhall

Ogbonnaya Ngwu

Amanuel Sahilu

Leonardo Villa-Forte

Sam Vučić

Zelai Xu

Tracy Lihui Zhang

Katerina Zhuravel

*Faculty Advisor:*

Harold Hodes



# Contents

---

Editors' Introduction	7
Restricting the Reactive Attitude: A Critique of Korsgaard's Moral Responsibility <i>Jordan Myers</i> <i>University of Pittsburgh</i>	9
Is Morally Meaningful Consent to Sex Work Possible? <i>Megan Wu</i> <i>Swarthmore College</i>	25
What to Make of Medical Nihilism? An Epistemological Defense of Optimism <i>Stanley Bowen Zhang</i> <i>University of Toronto</i>	39
Is Western Philosophy Irredeemably Racist? <i>Shariq Haidery</i> <i>St. Hugh's College, University of Oxford</i>	51
Stoic Love: A Senecan Defense of Friendship <i>Gabriel Sánchez Ainsa</i> <i>University of Chicago</i>	69
The Limits of Artistic Autonomy: The Case of Modernist Public Art <i>Alice Pessoa de Barros</i> <i>McGill University</i>	91



# Editors' Introduction

---

In some ways, 2020 has marked a resurgence for philosophy. As quarantine orders swept the globe, hot on the heels of a remarkably fast-spreading virus, our community of undergraduate philosophers reluctantly withdrew to the confines of our own homes. Forced to confront the heavy news of each passing day - some with more difficulty than others - we found ourselves with much to reflect upon, and even more time to write those reflections down.

Hence, this seventeenth edition of *Logos: the Cornell Undergraduate Journal of Philosophy* not only fulfills its annual duty of showcasing the best works of undergraduate philosophers around the world, it also serves as a record of our community's thoughts during a period of time unlike any other. But a global pandemic was not the only unprecedented situation for our editors: this year, we received a record-breaking 127 submissions from institutions around the world. From these 127 submissions, 46 advanced to a second round of review. From those 46 papers, 16 moved on to the final round, whence we selected our top six to be published.

We felt that these six papers were distinct in their clarity of language, development of argumentation and subject matter. Together, they comprise an impressive answer to the question: what were undergraduate philosophers thinking about in the year 2020? From moral responsibility to consent in sex work, medical nihilism to racism in Western philosophy, friendship to artistic autonomy, these papers each offered intriguing theses on the world around us. We hope you enjoy reading them as much as we did.

Yet, the true pleasure taken by our *Logos* editors was never contained within the bounds of these six papers. Each of the 127 papers we read rung a joyous reminder that our global community of undergraduate philosophers has continued to engage with each other, share their work, and find camaraderie in the shared pursuit for philosophical truths. To honor this dedication, the *Logos* editors took it upon themselves to attempt something just as unprecedented as the times they worked in: we provided custom feedback for every single paper we read. The papers we chose not to publish received these comments via email, while the six winning papers are each accompanied by a foreword written by our very own editors.

I cannot be more grateful for the dedication this year's editors have put into *Logos*. I would also like to thank the Sage School of Philosophy for their funding and support, especially Pamela Hanna and Dorothy Vanderbilt for their tireless commitment to helping our team in any way necessary. Finally, I would like to thank our faculty advisor Professor Harold Hodes. Thank you all, and please enjoy the thoughts and ideas that await.

*Bertrand Li*  
Editor-in-Chief





---

---

Restricting the Reactive  
Attitude: A Critique  
of Korsgaard's Moral  
Responsibility

---

---

*Jordan Myers*  
*University of Pittsburgh*

FOREWORD

*Editors Rebecca Sabin, Benjamin Mendenhall, and Ethan Kovnat*

What justification have we to hold people morally responsible for their actions? In “Restricting the Reactive Attitude: A Critique of Korsgaard’s Moral Responsibility,” Jordan Myers answers this question by arguing against Christine Korsgaard’s claim that we must regard one another as metaphysically free agents in all situations, which entails attributing moral culpability to others irrespective of outside forces that may have influenced their behavior. Myers not only describes potential exceptions to Korsgaard’s rule -- situations in which blame cannot be justified -- but also draws on the work of P.F. Strawson and Immanuel Kant to examine the issue of moral responsibility from multiple perspectives.. In doing so, Myers illustrates the nuances of ethical judgement and relates them to the problem of determinism and free will.

We selected this paper because it thoroughly analyzed Korsgaard and Strawson’s arguments, while maintaining sophisticated and clear prose. The paper presented the arguments in an engaging and understandable way that made it appealing to philosophers and non-philosophers alike. Furthermore, the paper’s opening paragraph really grabbed our attention -- we were hooked from the beginning!

This paper’s applicability to our lives and times cannot be overstated, as questions of responsibility and culpability will always be relevant and poignant. This paper is also impressive in how it presents metaethics accessibly and engagingly, which is a gift to those unfamiliar with a field that is too often criminally overlooked outside of academic philosophy.

I. INTRODUCTION

I would not like to be Christine Korsgaard’s friend – not because she isn’t presumably a lovely person and a deep thinker, but because being a close friend of hers may entail a level of commitment I am not prepared to engage in. Korsgaard holds a view that she ought always and totally hold her friends morally responsible, regardless of any deterministic considerations. While this would certainly be a welcome status quo for most of our engagements, I have good reason to believe her insistence on such a standard would end up straining and perhaps degrading our relationship as close friends. I believe it is not only permissible, but sometimes necessary, to withhold or suspend responsibility from others, even those whom we are close to. And it is for this reason which I playfully introduced my essay with its beginning line.

For most people, myself included, the practice of holding others morally responsible is an integral part of life. What I will aim to rebut in this essay, however, is Korsgaard’s argument on the constancy and necessity of this attitude.

Like Korsgaard, many people have the intuition that we should hold others morally responsible. Some philosophers maintain the existence of a metaphysical free will that grants a robust status to moral responsibility. Opposing this cohort are those who believe the universe consists of purely deterministic beings, and that this renders moral responsibility invalid. But there are also those who defy this dichotomy. P.F. Strawson finds questions of metaphysical determinism irrelevant to whether we can be considered free. Christine Korsgaard takes another view;

she argues that we can and must square the seemingly opposing standpoints and retain agentic responsibility in a determined world.

In this essay, I will first explain how Korsgaard's position – that we are compelled to always and totally hold individuals morally responsible – is largely shaped by Immanuel Kant's work on free will and his two viewpoints on persons. Then I will examine her thesis in her 1992 paper and 1996 book, both by the name *Creating the Kingdom of Ends*, and explicate what she has become logically committed to. After this analysis, I begin a critique of her thesis on its own grounds, then introduce P.F. Strawson's essay, "Freedom and Resentment," to inject a different perspective on moral responsibility. I explain Strawson's project of reconciling his optimist and pessimist and analyze his claims on the morality of reactive and objective attitudes. I also point out the areas of overlap and discord between Strawson and Korsgaard. Finally, I end with a critique of Korsgaard's absolutist views on agency and develop three cases in which I believe it is morally permissible to suspend Strawson's reactive attitude or Korsgaard's obligatory responsibility.

## II. UNDERSTANDING KORSGAARD'S ABSOLUTIST VIEW

Korsgaard derives her view on moral responsibility largely from Kantian philosophy, specifically Kant's work on the categorical imperative, freedom and determinism, and his transcendental idealism. Kant's second formulation of the categorical imperative famously states: "So act that you use humanity, whether in your own person or in the person of any other, always and at the same time as an end, never merely as a means."<sup>1</sup> From this, Korsgaard links holding someone morally responsible to treating her as an end in and of herself, and in the same way, to adhere to the second formulation of the Categorical Imperative. Korsgaard concludes that failing to hold others morally responsible is thus wrong because it is to break the second formulation; specifically, she argues that holding others responsible is how we treat them as ends-in-themselves, which is to say we respect their autonomy, liberty, and humanity as we do our own.

This raises two questions. One: Is Korsgaard committed to the view that we must always hold people responsible without exception? And two: Is this view problematic or compelling? I will argue that Korsgaard must be logically committed to the view proposed, even though she herself backs away from this conclusion, and that her answer is problematic because she fails to define why certain cases of action warrant exclusion from responsibility. The first question can be examined in light of how Korsgaard and Kant view subjective free will and how they view two disparate stances on the human condition. The second question will be answered through a critique of Korsgaard's position on its own terms and a Strawsonian analysis of the critiques raised against Korsgaard.

Korsgaard first argues for the absolute, exceptionless status of moral responsibility. She does so by supposing that when we act, from a subjective frame of reference, we must "...regard ourselves as active beings..." that can authoritatively deliberate, not some mere byproduct of our circumstance.<sup>2</sup> She

---

<sup>1</sup> Kant, Immanuel. *Groundwork of the Metaphysics of Morals* (Edited by M.J. Gregor. Cambridge, U.K.: Cambridge University Press, 1998), [G 4:429].

<sup>2</sup> Korsgaard, Christine M. "Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations." *Philosophical Perspectives* 6 (1992): 317.

claims that we simply must and always do feel like the total author of our actions, and from this, she claims we always view *ourselves* as morally responsible, at least in moments when making decisions.<sup>3</sup> She follows Kant in claiming that when we act, we cannot help but view ourselves “under the idea of freedom... We cannot conceive of a reason which consciously responds to a bidding from the outside with respect to its judgements.”<sup>4</sup>

Korsgaard elaborates on Kant’s views: “The point is not that you must believe you are free, but that you must choose as if you were free. It is important to see that this is quite consistent with believing yourself to be fully determined.”<sup>5</sup> She exemplifies this by way of a thought experiment in which an electronic device is implanted in your brain. The device does not circumvent your thought process, but rather injects thoughts into the flow in consciousness seamlessly, as if they were your own. This way, the thoughts manifest themselves in such a way that every thought and subsequent action *subjectively feel* as they normally would. Korsgaard adds:

...perhaps you get up and decide to spend the morning working. You no sooner make the decision than it occurs to you that it must have been programmed. We may imagine that in a spirit of rebellion you then decide to skip work and go shopping. And then it occurs to you that *that* must have been programmed.<sup>6</sup>

This idea is meant to persuade the reader that we are helpless but to “regard ourselves as having free will” because even in a scenario where we know we are programmed, we cannot function unless we practically accept that we still must act as agentic beings.

Kant elaborates that if freedom of the will is presupposed, morality along with its principles follows from it.<sup>7</sup> Curiously, Kant also claims later in *The Groundwork* that morality itself provides grounds from which free will can be deduced.<sup>8</sup> Whether these two statements constitute a contradiction, change of mind, or complementary views does not matter for this discussion; what is clear is that Kant – and subsequently Korsgaard – view free will and morality as closely tied together, dependent in some way on each other. This brings us closer to understanding how Korsgaard arrives at her own interpretation of the second formulation of the Categorical Imperative.

Korsgaard has now argued that we always treat ourselves with agentic responsibility, and that it is only with others that we tend to ‘explain away’ their actions as caused by factors outside their control. But Korsgaard contends that since we do not take this genealogical view of our close friends – that is to say, we hold them responsible as we would ourselves – we can widen this circle to encompass all rational beings, since all friends were once strangers.<sup>9</sup> At this point, Korsgaard is logically committed to granting that others must view themselves as moral agents like we view ourselves.

<sup>3</sup> Korsgaard, “Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations,” 319.

<sup>4</sup> Kant, *Groundwork of the Metaphysics of Morals*, [G 4:448].

<sup>5</sup> Korsgaard, Christine M. *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996), 162.

<sup>6</sup> Korsgaard, *Creating the Kingdom of Ends*, 162-163.

<sup>7</sup> Kant, *Groundwork of the Metaphysics of Morals*, [G 4:447].

<sup>8</sup> Kant, *Groundwork of the Metaphysics of Morals*, [G 4:454].

<sup>9</sup> Korsgaard, “Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations,” 319-320.

The second part of Korsgaard's argument for moral responsibility stems from her discussion of two standpoints from which we may view a human being: the noumenal and phenomenal. The phenomenal standpoint is concerned with sensible things, or things as they appear to us, whereas the noumenal realm concerns itself with the laws that ground the sensible world.<sup>10</sup> A person has a foot in each realm, so to speak. He can be viewed from a phenomenological view that paints him as a sensible, determined object of prior causation, or from a noumenal view as a rational, agentic causer of his own actions. According to Kant and Korsgaard, neither of these views tread on the other's territory. The noumenal and phenomenal share in non-overlapping magisteria, and can thus both be accurate while viewing the same person very differently.<sup>11</sup>

How do these standpoints influence responsibility? The key here is that Kant does not view free will in strong metaphysical terms. Instead, recall that Kant argues free will is something we must adopt from the project of ethics itself<sup>12</sup> or because, if free will is presupposed, it creates the possibility of morality along with its principles.<sup>13</sup> Because of this commitment, Kant is summarized nicely by Korsgaard as stating "The person needs the belief as a condition for obedience to the moral law, and it is this, combined with the categorical nature of the law, that justifies the belief. Although the basis of the beliefs is theoretical in form – the will is free, there is a God – their basis and function are practical. As Kant says... the postulates play no theoretical or explanatory role whatsoever."<sup>14</sup>

Both the noumenal standpoint – that human beings are agents who are in causal command of their actions – and the phenomenal standpoint – that human beings are vitiated by causality outside their control – are 'true' from their respective viewpoints. Korsgaard accepts this seeming paradox:

Insofar as we view our actions as phenomena we must view them as causally determined, but not necessarily as determined by mere desires and inclinations. We can still view them [our actions] as determined by moral thoughts and moral aspirations; only from this point of view, those must themselves be viewed as determined in us.<sup>15</sup>

Here Korsgaard foists the moral culpability on us, not our circumstances, even though they are still applicable. She accepts that factors outside our control explain our actions, but they are still *our* moral actions, and so it is we who must be held responsible. Blaming circumstance or luck would bypass what it means to hold someone accountable as an autonomous individual.

It is through this route that Korsgaard arrives at her conclusion: that we must always hold individuals morally responsible. The act of holding responsible is a result of accepting and granting that individuals have free will, not on theoretical grounds, but on simpler, more interpersonally pragmatic grounds; it is this act of holding responsible that keeps morality itself alive and well on Korsgaard's view. This is a practical law that has a categorical nature; thus, her conviction to hold

---

<sup>10</sup> Kant, *Groundwork of the Metaphysics of Morals*, [G 4:453].

<sup>11</sup> Korsgaard, *Creating the Kingdom of Ends*, 168-169.

<sup>12</sup> Kant, *Groundwork of the Metaphysics of Morals*, [G 4:454].

<sup>13</sup> Kant, *Groundwork of the Metaphysics of Morals*, [G 4:447].

<sup>14</sup> Korsgaard, *Creating the Kingdom of Ends*, 172.

<sup>15</sup> Korsgaard, "Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations," 322-323.

responsible logically admits of no exceptions! Additionally, Korsgaard claims to not value one standpoint more than the other, but then adds that "...if either is privileged, it is the practical, because, according to Kant, 'every interest is ultimately practical' (C2 121)."<sup>16</sup>

I understand these views as logically committed to always *and without exception* holding individuals responsible. Korsgaard first maintained that others must view themselves as moral agents as we view ourselves. Then she argued that regardless of deterministic causes, moral actions are still those of their agents. She postulated that we do not hold people responsible because of the reactions it would cause in them – not for consequentialist reasons – but rather because it is something we *do* as we see each other as moral agents.<sup>17</sup> So then, if we cannot help but view ourselves as responsible moral agents, and we grant that others view themselves this way, we must always view others in the light we view ourselves, and thus we must always hold others responsible. Even though there are explanatory stories about how we developed, since they do not apply to our subjective deliberations, and because those deliberations are shared by all rational beings, there is no ground to excuse or waive the responsibility of others.

Korsgaard strikes a unique pose in the philosophical community on questions of responsibility. As introduced by Aristotle (from the Stanford Encyclopedia of Philosophy), there are at least two basic ways of viewing this debate. Praise or blame may be warranted because of the reactions or consequences it would produce in the other person, such as a change of behavior.<sup>18</sup> This *consequentialist view* of moral responsibility is explicitly disavowed by Korsgaard – after all, she does take a Kantian line on responsibility and would thus be opposed to a consequentialist view. The other is often referred to as the *merit-based view* of responsibility; praise or blame is warranted if such a reaction is deserved or found to be appropriate given the agent's character or actions.<sup>19</sup> Korsgaard does not exactly fit neatly into this camp either, however. She seems to agree with the general idea of the merit-based view, but concludes that the conditions are *always* such that praise or blame is appropriate, cancelling the need for the view's conditional clause. We might say that Korsgaard holds to an extremist version of the merit-based view, a view of extreme responsibility, not based on a view of the universe as indeterminate, but from the standpoint of requiring responsibility for moral, interpersonal life itself.

### III. CRITIQUING KORSGAARD'S ABSOLUTIST VIEW OF RESPONSIBILITY

The above exegesis introduces my first critique of Korsgaard; she is incorrect in denying our capacity to experience ourselves as deterministically caused – not as agents who act, but as objects who are acted upon. My second critique consists in illuminating a contradiction between her descriptive claims of subjective agency and her normative views on what we ought to do with this self-knowledge.

Firstly, I want to raise an introspective critique of Korsgaard; she claims that we cannot help but act in an agentic manner, but the firsthand experience

<sup>16</sup> Korsgaard, *Creating the Kingdom of Ends*, 173.

<sup>17</sup> Korsgaard, "Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations," 305.

<sup>18</sup> Talbert, Matthew. "Moral Responsibility," *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition), Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/win2019/entries/moral-responsibility/>.

<sup>19</sup> Talbert, "Moral Responsibility."

of myself and others – which is Korsgaard’s basis of argumentation as well – demonstrates the contrary. In the practice of Buddhist Vipassana meditation, for instance, practitioners can glimpse prolonged moments of clarity wherein thoughts cease to appear as one’s own, but rather arise from an opaque darkness of subjective mystery. This is not a result of distracting oneself, however; it is a result of a conscious effort to more closely examine the mind. Such an experience is direct evidence to the contrary of Korsgaard’s view espoused in “Creating the Kingdom of Ends.”<sup>20</sup>

Korsgaard, though, denies that such moments of mental perspicacity exist in principle. She may rebut this critique and clarify that her view of agency is only applicable when *acting*, and the practice of meditation, or simply introspecting on how the mind works, is passive in nature and thus inapplicable. But she never argued for any strong separation between passive and active moments in terms of how the mind operates; the boundary between the two states of mind seems erroneous when one considers that introspection itself is indeed an active practice of simply paying attention to one’s own mind.

The second aspect of my critique is to unearth a lurking contradiction between Korsgaard’s descriptive state of what is possible and her normative claims of what is moral – I should add that this critique can stand alone from the previous one. They are related, but not dependent. As previously examined, Korsgaard claims that we *cannot* see ourselves as determined in the present, only in the past or future. I have just provided a direct challenge to this description of what is subjectively possible. But she then adds that one *ought not* normatively see oneself as determined in the future because that would reduce the only responsible person to oneself in the current moment: “... You cannot restrict the concepts of freedom and responsibility to yourself in the context of deliberative choice. If you did, you would think that the only free agent in the world is merit-now.”<sup>21</sup> She adds that “... unless you regard others and your future self as moral agents, there will be no content to your duties at all, for all duties (according to Kant) are owed either to other persons or to the enduring self.”<sup>22</sup> Clearly, we must, on her account, hold our futures selves responsible!

But herein lies the paradox: Korsgaard explicitly admits that I can view my future self as *a thing* deterministically caused. For instance, when actively planning the rest of my day, I might choose to write this paper in the university’s public library if I know I will be more motivated to work there, as others will be judging me. Here I view myself in deterministic terms: “*In planning my afternoon, I know that if I go to the library and work around others, I’ll get more work done. If I go home and work from my couch, I’ll get less work done. Therefore, I’ll go to the library to work.*” In the same thought, I might refuse myself chocolates at the grocery store if I know I will later lack the willpower to consume them in moderation. I may act now to automatically divert a portion of my paycheck to a locked savings account because I know that if I leave open the possibility, I may frivolously spend more than is wise. I might say no to dinner if I have an early start tomorrow, because I know that if I go to dinner, I will consume more wine than is advisable and will fail to wake up on time or in the proper condition tomorrow morning. These are instances where I act as an agent – I am making the choice –

<sup>20</sup> Korsgaard, “Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations,” 319.

<sup>21</sup> Korsgaard, “Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations,” 320.

<sup>22</sup> Korsgaard, “Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations,” 320.

but also as a determined outcome – if I do action X now, I will be helpless but to do Y later.

We can intelligibly speak about cases like the ones above, wherein I view my future self in deterministic terms, but this act is forbidden by Korsgaard's categorical prohibition – it would be to violate the second formulation. But to admit that cases like these are *intelligible* is to admit that that they may be – and often are – *true* of ourselves! In admitting their truth but forbidding their use, Korsgaard weakens her own claim about our experiential knowledge. She forces us to ignore, to cast our gaze from this self-knowledge – the very thing she built her case of subjective agency upon! She forces us to pretend, to lie to ourselves, and imagine deterministic facts are not and will not be true of us. This would resemble denying that I have been unable to dunk a basketball because I am too short – likewise, I have not been able to go to dinner without ordering wine, work from the couch as efficiently as from the library, or buy chocolates now for moderate consumption later.

To admit that deterministic causalities were true of us in the past is to admit that they were once true of a present-tense self, and that if they are true in the future, they will also be true of a present-tense self. But admitting that a deterministic view is true of our present self is the very thing which she denied was descriptively possible, while also incompatibly laying the grounds to conclude it *is* descriptively possible but morally forbidden!

I may look back at my patterns of behavior – perhaps caused by any number of deterministic details: past trauma, brain damage, transformative experiences, parental genetics – and use that understanding to pragmatically shape my future under the belief that I will continue to be ruled by these deterministic variables. In the same way that I cannot change my height or lactose intolerance, I may be unable to bootstrap my motivation, courage, or anxiety in certain circumstances. Merely asserting that I can change my motivation to work from my couch is in some sense akin to asserting that I can override my intolerance to lactose through sheer willpower. It is this lurking contradiction that I believe illuminates Korsgaard's untenable position – she seems to be having her descriptive cake and enjoying the normative slices as well.

I believe I have successfully raised two objections to Korsgaard's descriptive claim on how I can subjectively view myself; I now must demonstrate that viewing ourselves or others as deterministically caused can ever be morally permissible, now that I have shown it is at least descriptively possible. Recall the style of my critique: I aim to show that Korsgaard has argued for grounding that logically entails a commitment to the ubiquitous application of moral responsibility, but first challenge her grounding views in what I have just raised. I will then go on to show there are normative impossibilities to Korsgaard's view, and these cases are ones that Korsgaard herself attempts to concede, but without any proper theoretical grounds for doing so.

#### IV. OBJECTIVE AND REACTIVE ATTITUDES

Injecting P.F. Strawson's view, explained in *Freedom and Resentment*, may help elucidate the differences between myself and Korsgaard. Strawson views himself as one who “does not know what the thesis of determinism is.”<sup>23</sup> He

---

<sup>23</sup> Strawson, Peter. “Freedom and Resentment,” *In Perspectives on Moral Responsibility*, edited by John Martin Fischer and Mark Ravizza (Ithaca and London: Cornell University Press, 1993), 45.



sees questions of metaphysical free will and determinism as unrelated and inconsequential to a discussion of moral responsibility. Strawson is not concerned with the question of some ultimate, perfectly uncontaminated sense of free will; rather, his explicit goal is to reconcile the pessimist – those who think that if determinism is true, moral responsibility makes no sense – and the optimist – those who deny that moral responsibility loses any power or appropriateness if determinism is true.<sup>24</sup> Strawson’s pessimist can also be understood as holding an *incompatibilist* view of responsibility, where moral responsibility is incompatible with a deterministic universe.<sup>25</sup> The contrapositive is that, for moral responsibility to make any sense, the universe and the creatures in it must be indeterminate or unconstrained in the relevant manner. His optimist, on the other hand, can be thought of as a *compatibilist*, someone who believes individuals can still be held responsible even if their actions or personhood is fully determined by forces and causes outside themselves. Strawson aims to *dissolve*, rather than *resolve*, the conflict between these two positions.<sup>26</sup>

Strawson postulates that we take two different types of attitudes towards others: reactive attitudes and objective attitudes. Reactive attitudes are those we take when seeing another person as agentic and as an appropriate target of our reactions. Reactive attitudes can be positive or negative; Strawson speaks of resentment and gratitude as an opposing pair, but there are countless others as well. “...There is a whole continuum of reactive attitude and feeling stretching on both sides of these and – the comfortable area – in between them.”<sup>27</sup> Simply put, reactive attitudes are those which are integral to how we interact as people: forgiving, blaming, resenting, being grateful towards, holding in high esteem, empathizing with, being embarrassed in front of, and the like. One would not find it reasonable or appropriate, on this view, to hold a reactive attitude towards an inanimate object, like a car that has broken down or a rocket that has successfully launched.

Objective attitudes require the suspension of this interpersonal involvement. In Cheshire Calhoun’s paper “Changing One’s Heart,” this can take the form of separating the action from the culpable person, viewing the action with an objective attitude and the person still with a reactive attitude.<sup>28</sup> Or more powerfully, objective attitudes can take the person in question and degrade him from an agentic individual with which we may have a relationship and transform him into an object of deliberation, something in the situation to be accounted for deterministically.<sup>29</sup> The second sense of the objective attitude holds more power and will be the main focus of this analysis.

The concept of “being responsible” is used here in a specific way. A physicist may say something like “*The sun’s gravity is responsible for Earth’s orbit around it*” or an accountant may state “*The sales department is responsible for the bloated expenditures this quarter.*” Neither of these definitions – loosely defining *responsibility as attributability* and *ledger responsibility* – get at the core of moral responsibility.<sup>30</sup> Instead, for this discussion, moral responsibility

<sup>24</sup> Strawson, “Freedom and Resentment,” 45-46.

<sup>25</sup> Talbert, “Moral Responsibility.”

<sup>26</sup> Talbert, “Moral Responsibility.”

<sup>27</sup> Strawson, “Freedom and Resentment,” 50.

<sup>28</sup> Calhoun, Cheshire. “Changing One’s Heart.” *Ethics* 103, no. 1 (1992): 78.

<sup>29</sup> Strawson, “Freedom and Resentment,” 52.

<sup>30</sup> Talbert, “Moral Responsibility.”

refers to the appropriate application of Strawson's reactive attitudes. Korsgaard's responsibility is an example of a reactive attitude; it is more encompassing than simply being able to empirically know who caused certain outcomes. She explains responsibility as being prepared to

...exchange lawless individual activity for reciprocity... to accept promises, offer confidences, exchange vows, cooperate on a project, enter a social contract, have a conversation, make love, be friends, or get married... to risk your happiness or success on the hope that she will turn out to be human.<sup>31</sup>

Korsgaard would agree with Strawson that holding the reactive attitude is something we *do* as individuals engaged in relationships, and that it is central to our daily lives. She would also agree with Strawson that the key focus is not on a metaphysical, abstracted, ultimate sense of uncaused free will; rather, the free will discussed is one of agency, responsibility, and practical interpersonal exchanges. They both claim that "In everyday personal interaction, we cannot get on without the concept of responsibility."<sup>32</sup>

Strawson merely claims that sometimes we *do* take the reactive attitude and that other times we *do* take the objective attitude. But he is not clear on what criteria *should* determine which attitude we take in what circumstances.<sup>33</sup> In fact, this is a deliberate choice on Strawson's part; he says that someone like myself has been all too busy overintellectualizing the problem. The act of holding responsible, on his account, "... neither calls for nor permits, an external 'rational' justification."<sup>34</sup> Korsgaard makes it quite clear that taking the objective attitude is *always* wrong, but Strawson's view is more difficult to parse. He explicitly states we simply cannot give up the reactive attitude altogether because it is too ingrained in us, and that even if, theoretically, determinism was true, this would not bear on the practical question of reactivity.<sup>35</sup> But one can also understand Strawson as saying that for someone who is free, holding the reactive attitude towards him is not a descriptive claim about him, but an indication that we *do* hold the reactive attitude towards that person.<sup>36</sup>

I find this latter version of Strawson to be a rather dull and uninformative one. It merely states that saying someone is free, holding them responsible, and having reactive attitudes towards him are all part of the same response. It is a sort of radical redefining that does not arrive at a convictive conclusion. I contend that it lacks philosophical rigor and explanatory power, as well as informative power; it merely describes some of our actions and avoids engaging with the messiest parts. On this view, we have nowhere to stand when asking ourselves or others if an attitude is in fact appropriate! The more important question strikes me as: *under what conditions is the objective attitude permissible or even obligatory?* This is the question I want to unpack and wrestle with, and doing so requires a critique of Korsgaard and Strawson in a way that creates room for a new view to emerge.

---

<sup>31</sup> Korsgaard, "Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations," 306.

<sup>32</sup> Korsgaard, "Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations," 312.

<sup>33</sup> Strawson, "Freedom and Resentment," 59-60, 63.

<sup>34</sup> Strawson, "Freedom and Resentment," 64.

<sup>35</sup> Strawson, "Freedom and Resentment," 59-60.

<sup>36</sup> Strawson, "Freedom and Resentment," 66.

## V. WHEN IS THE OBJECTIVE ATTITUDE APPROPRIATE?

There is a point in “Creating the Kingdom of Ends” wherein Korsgaard says something interestingly contradictory. She appears to seemingly ‘slip-up’ and admit that she has partially sympathetic views to Strawson and myself. She presents a seeming concession, citing Kant: “... it may be reasonable, when we are deciding whether and when to hold people responsible, to take into account such things as upbringing and education.”<sup>37</sup> But earlier Korsgaard was adamant that we *ought not ever* forgo responsibility for considerations of determinism. She also supportively cites Kant, saying “... when someone makes an error, [it is right] not just to deem him stupid but to try to determine how the mistaken view could have seemed reasonable to him.”<sup>38</sup> This project begins to creep closer towards an objective attitude, something Korsgaard categorically disavowed.

She adds that sometimes we simply “...laugh them [errors] off as the result of a transitory emotion or exhaustion.”<sup>39</sup> This option is supposed to supplement the urge to either write off the person as stupid or unteachable, or making his/her error understandable. This is an attempt by Korsgaard to make room for absolving responsibility while still not undermining the individual’s integrity. But this attempt seems hard to square with what she previously wrote about accepting a deterministic paradox. Recall that our actions are “...determined by moral thoughts and moral aspirations... those must themselves be viewed as determined in us.”<sup>40</sup>

Here arises a contradiction; Korsgaard’s example of laughing off a transitory state of consciousness conflicts with her initial position, that regardless of how our condition developed, it is still our *responsibility* to act morally. She does not provide a *reason* why certain cases like this should be exempt! It appears Korsgaard capitulated to her own argument for moral agency, then suddenly realized that it was an impossible standard to adhere to. She then trotted out exceptions by fiat that do not explain *why* they warrant exemption.

But if Korsgaard is willing to create cases lacking responsibility without firm bounds, there is nothing that determines which of these cases are allowed, and therefore any case would be open to a removal of responsibility. Her project then becomes one of brute intuitions and hand-waving, not a circumspect theory of a Kantian philosopher. This gives rise to why I argue Korsgaard ties herself to a problematic view, in a way that she herself anticipates. This is also Strawson’s error; he creates an account of attitudes that does not provide simplicity, power, or explanation – qualities that philosophers like Shelly Kagan demand of a moral theory.

The question then remains: When is a suspension of the reactive attitude appropriate? First, I will concede that, as per Korsgaard and Strawson, the answer cannot be ‘always.’ A life vitiated by the objective attitude would be a lonely existence. The answer likewise cannot be whenever one pleases. Critiquing Korsgaard and Strawson within the bounds of their own terms, I argue there are three cases in which it is morally permissible to hold the objective attitude.

---

<sup>37</sup> Korsgaard, “Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations,” 323.

<sup>38</sup> Korsgaard, “Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations,” 324.

<sup>39</sup> Korsgaard, “Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations,” 324.

<sup>40</sup> Korsgaard, “Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations,” 323.

The first case arises when one holds the objective attitude towards oneself in a moment of deliberation or reflection. As I previously discussed, viewing oneself through an objective lens is not only a largely accurate account, it can be extremely useful in practical planning. Knowing my personal quirks, habits, and tendencies – also recognizing my inability to escape them – and taking them into account when planning is clearly beneficial. The former is an example of deterministic reflection, the latter of deliberation. I can retrospectively view myself in deterministic terms to discover what about me produced certain outcomes in my life and prospectively view myself deterministically in order to create a better future.

Taking the objective attitude towards oneself during agentic deliberation also negates Korsgaard's concern about degrading interpersonal relationships because this objective view is taken towards *myself*. Kant, however, may raise a concern that the objective attitude fails to respect *oneself* as an autonomous individual. I do not think this concern is valid either. Recall that viewing oneself in deterministic terms still occurs within the bounds of *agentic* reasoning. I consent to viewing myself deterministically.

Korsgaard might spring to Kant's rescue, accusing me of absolving myself of *any* reason to hold myself responsible. But to this remark I may call on Korsgaard herself; recall her claim that no matter what, we still must take agentic action, even if we know ourselves to be fully determined. She says that even acknowledging the thesis of determinism "... can only prevent you from making any decision. In order to *do* anything, you must simply ignore the fact that you are programmed, and decide what to do – just as if you were free."<sup>41</sup> So, by her own account, it is impossible for me to fully collapse into viewing myself in exclusively objective terms, which I am happy to agree would be a personal death of sorts. Thus, when I do view myself deterministically, which is to say objectively, I can only do so within the larger context of agent-centered deliberation, which alleviates the slippery slope into pure objectivism.

The second appropriate application of the objective attitude comes when interpersonal relations have been degraded *by the same person whom you are tempted to hold the reactive attitude towards*. It is crucial that the relationship be compromised by the *other* person in question, not you. Adhering to this constraint mitigates Korsgaard's concern with dehumanization. I agree with Korsgaard that responsibility comes along with the status of personhood, and deciding to rescind this status is a moral peril.<sup>42</sup> Thus, the fault must be with your friend, colleague, family member, or other. I am aware that this remark can be construed as selfishly excusatory, but I am speaking in precise terms of what I believe to be the morally correct view, not whether I am likely to abuse it or deceive myself.

If your boss or neighbor, for instance, routinely lacks the capacity for reciprocal interactions, you cannot be morally obligated to engage with him under the duty of holding reactive attitudes. Such a project is not functionally possible, let alone worthwhile. Imagine the details of such an obligation: you would be morally held to becoming angry with someone who could not understand your anger; you would be forced to reason with someone who *could not* or *would not* accept reason on its own grounds; you would be forced to reconcile your differences with someone who was not interested in continuing the relationship. These situations

---

<sup>41</sup> Korsgaard, *Creating the Kingdom of Ends*, 163.

<sup>42</sup> Korsgaard, "Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations," 314.

can be painted ad nauseum and illuminate the impracticality of an unwavering grasp on reactive attitudes.

The third case is one in which the reactive attitude is inappropriate due to an intrusion of an overwhelmingly causal detail, one that demands a suspension of the reactive attitude. This is also the only case in which I will argue the reactive attitude is impermissible. The term ‘overwhelmingly causal detail’ describes something about a situation or constitution of an individual that renders him an inappropriate target of the reactive attitudes. This key detail must be something such that, if it was not present, the reactive attitude *would* be appropriate.

Perhaps the way to explain this exception is best done by returning to Korsgaard’s thought experiment in *Creating the Kingdom of Ends*. Recall that she used the idea of a machine implanted in your head which fully determines your every move for that day. She claimed that, as a result of considering this thought experiment, it is impossible to subjectively differentiate between self-created actions or those which fully arose from the machine.

Consider a variation of Korsgaard’s thought experiment. Here the machine only induces half of your actions and leaves the other half totally unaffected. Crucially, the machine does not indicate to its host or the outside world which actions were induced, and which were not. Imagine that your friend had this device installed for one day as part of a research study, and during this day, your friend attacks and kills a stranger for crassly insulting his physical appearance. Obviously, the act of killing a man over mere insults is morally abhorrent, but Korsgaard’s claim that a determined action would feel identical to an agentic one does not clarify this situation.

What we want to know – *need to know* – is whether the act of murder was induced by the machine, or if it was the act of a normally agentic individual. Capturing this information would be the top priority of any sane person; I cannot imagine someone who would not care to learn of it. Imagine what it would take to deny that such information was relevant to how the case proceeded. Such a person would be committed to a worldview that did not include even the most transparent details of a situation. This person would be the moral equivalent of a totalitarian dictator – someone who sentenced his subjects without needing to hear any details of the crime in question.

Because of her previous arguments, Korsgaard is committed to denying that deterministic details like this matter at all and in principle. I raised a variant of her thought experiment as a *reductio ad absurdum* to her view; Korsgaard must either abandon the arguments which tie her to this conclusion, and admit that there are clearly cases wherein the constitution of the person was compromised beyond reasonable standards of responsibility, or she must accept the absurdity of her view. If she chose the latter, I suppose there would be nothing more to say in hopes of persuasion. I would be alienated from her moral worldview altogether, and view her in objective terms as someone who could not see what must be clear. I could no longer understand what she meant by responsible, and moreover, I do not believe *she* could understand what she meant by responsible. The lines of causal connections would be so blurred as to render them useless. If the source of my actions can be perfectly altered without my consent being required, and I am no less responsible for that outcome, then moral responsibility ceases to mean anything of substance. The very reason we maintain concepts of responsibility or exoneration are because we do and should recognize that cases

exist on both sides of the divide. Korsgaard's view is too simplistic; it betrays too low a resolution for scrupulous moral thinking.

I should briefly stave off Korsgaard's likely objection that I have presented a legal case and equivocated it with a personal one. Korsgaard takes a hard line of separation between legal and interpersonal cases, but I want to apply pressure to this distinction in support of my argument. I find it clear that my example of the murderer is *both* a legal *and* interpersonal case. Clearly, the case will be settled at the legal level – with a judge, jury, and perhaps executioner – but it must be clear that there is more to the story. The murder will affect people and their relationships, their lives, and their loved ones. When our society settles matters legally, we are in a constant process of importing and exporting personal concerns into the process. We want to know who committed the crime, how it affected the victim and his family, and perhaps what we can do to help. The crime is a legal dispute *and* a personal one! This paper does not have the scope to encompass a full discussion of the issue, but I hope to have raised a brief but adequate rebuttal to Korsgaard's likely objection.

The reactive attitude is clearly in need of finer calibration. As Thomas Nagel explicated in his essay, "Moral Luck," our attitude towards wrongdoers is often clouded by the arbitrary or incidental outcomes of identical actions.<sup>43</sup> If two of my friends, Brian and Adam, are both driving home from work, and they both send a text while driving for the same amount of time and on the same roadway, but Brian hits and kills a child and Adam arrives home without incident, we will hold extreme reactive attitudes towards Brian while leaving Adam unreprimed. Thus, an identical action can produce radically different results and radically difference reactive attitudes.

I raise this point to show that Strawson's reactive attitudes, and Korsgaard's insistence on the absolute application of one in particular, are not crafted to the standard that academic philosophy demands. Random moral luck does not create fair conditions for blame and resentment, and I would imagine a Kantian philosopher like Korsgaard to be sensitive to respecting the rights of others to be treated fairly.

## VI. CONCLUSION

In this paper, I have raised two simple objections to Korsgaard's view on subjective agency and used these to begin a larger critique of both Korsgaard and Strawson. I do largely agree with much of both projects, however. I agree with Strawson that to hold someone responsible is something we do when we view him as an a-metaphysically free actor, and as per Korsgaard, this attitude is required for viewing someone as a person with whom we might hope to maintain a relationship. I differ from Korsgaard in arguing that under certain conditions, the suspension of such reactive attitudes is morally permissible or even obligatory. This claim is also bolder than Strawson's descriptive state of attitudes. I point out a contradiction within Korsgaard's work that leads to a presentation of three cases in which I believe a suspension of the reactive attitude is morally permissible or obligatory.

The first permissible case is when I hold the objective attitude towards myself in moments of active deliberation or reflection. Here I view myself as a determined

---

<sup>43</sup> Nagel, Thomas. *Mortal Questions* (Cambridge: Cambridge University Press, 2012), 24-38.

thing to be examined or taken into account. The second case of moral permissibility occurs when I suspend my reactive attitudes towards someone who will not or cannot engage interpersonally in a way that warrants reactive attitudes. This person has degraded my relations with him to the point of non-collaboration, and has thus brought the objective attitude on himself. The third case is one where deterministic circumstances demand the objective attitude, cases in which a key variable negates even Korsgaard's grounds for practical responsibility. I raised a variation of Korsgaard's thought experiment in support of this third case, and concluded with a supportive example from Thomas Nagel of how our reactive attitudes can be easily misdirected. I find great value in what Korsgaard and Strawson assert, but push for a change in Korsgaard's absolutism and a bolder stance than Strawson's. Holding people in our lives responsible is central to interpersonal relationships, but it is not morally responsible for us to default to the reactive attitude.

#### BIBLIOGRAPHY

Calhoun, Cheshire. "Changing One's Heart." *Ethics* 103, no. 1 (1992): 76-96.

Kant, Immanuel. *Groundwork of the Metaphysics of Morals*. Edited by M.J. Gregor. Cambridge, U.K.: Cambridge University Press, 1998.

Korsgaard, Christine M. "Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations." *Philosophical Perspectives* 6 (1992): 305-332.

Korsgaard, Christine M. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press, 1996.

Nagel, Thomas. *Mortal Questions*. Cambridge: Cambridge University Press, 2012.

Strawson, Peter. "Freedom and Resentment." In *Perspectives on Moral Responsibility*, edited by John Martin Fischer and Mark Ravizza, 45-66. Ithaca and London: Cornell University Press, 1993.

Talbert, Matthew. "Moral Responsibility." *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition), Edward N. Zalta (ed.), <<https://plato.stanford.edu/archives/win2019/entries/moral-responsibility/>>.





---

---

# Is Morally Meaningful Consent to Sex Work Possible?

---

---

*Megan Wu*  
*Swarthmore College*

## FOREWORD

*Editors Fabio Cabrera, Tracy Libui Zhang*

What grounds ethical sex and how does consent tie to it? When, if ever, does sex work become morally impermissible? How and why is our current social and legal conception of consent not only exceptionally inadequate but sexually exploitative?

In *Is Morally Meaningful Consent to Sex Work Possible? In Defense of Enthusiasm as a Necessary Criterion for Sexual Consent*, Megan Wu provides a multinational historical account of sex work and presents the inadequacies of established notions on sexual consent. After outlining the relationship between the application of these ideas and sexual exploitation, Wu provides a vitally missing component to both of these models: the enthusiasm criterion. Wu not only robustly fortifies this principle, but provides an ethical apparatus for its application, producing a deeply provocative paper. The profound clarity of Wu's ethical analysis gives power to the author's case that meaningful consent to sex is not necessitated by the mere existence of free choice (Contractarian Model of Consent) or potential alternative forms of employment (Feminist Model of Consent), forcing us to reconsider our legal and social approach to sex work and sexual consent. With Wu finally astutely discerning that the current sex industry at large fundamentally lacks consent, we are left to ponder the heavy ethical question: can sex work be morally permissible?

## I. INTRODUCTION

Despite being among the oldest human professions, sex work is highly stigmatized in modern liberal democracies like the United States. Sex workers face criminal penalties everywhere in the United States, except for some rural counties in Nevada, which makes it difficult for sex workers to leave the industry or redress abuse they experience in the line of work; this has spurred a movement calling for decriminalization of sex work<sup>44</sup>. Given the increased focus in recent years on consent as a necessary and sufficient condition for morally permissible sex, the accompanying push for normalizing sex work and treating it like other forms of service labor merits thorough consideration.

In this paper, I define sex work as the exchange of sexual services for money. Because sex work requires the participation of at least two parties, it "should be understood to include the combined activity of all these parties"<sup>45</sup>, not just the activities of the sex worker. To be explicit, any criticisms of sex work that I make are therefore not criticisms of the behavior of sex workers, but of the existence of the sex work industry, the existence of demand for the sex work industry, and the practices that occur in it. Importantly, this definition also does not include sex trafficking, where persons, not their services, are bought and sold against their will. Nor does it include underage sexual labor of any sort.

---

<sup>44</sup> Anna North, "Sex Workers' Fight for Decriminalization, Explained," Vox, August 2, 2019, <https://www.vox.com/2019/8/2/20692327/sex-work-decriminalization-prostitution-new-york-dc>.

<sup>45</sup> Scott A. Anderson, "Prostitution and Sexual Autonomy: Making Sense of the Prohibition of Prostitution," *Ethics* 112, no. 4 (July 2002): 748–80, <https://doi.org/10.1086/339672>, 248.

Another important note to make here concerns the term “sex work” itself. On one hand, usage the term “sex work” may normalize the idea that sex work is analogous or comparable to other, non-sexual forms of “work”<sup>46</sup>. On the other hand, synonymous terms like “prostitution” also carry the negative connotation and social stigma directed towards sex workers. In this paper, I choose to use the term “sex work” over prostitution, not to normalize the idea that sex work is beyond criticism because it is a form of labor, but because all labor and forms of work are themselves not beyond criticism. As Banyard, quoting philosopher Debra Satz, writes: “All markets depend for their operation on background property rules and a complex of social, cultural, and legal institutions”<sup>47</sup>.

The notion that “it is possible to commodify consent”<sup>48</sup> has a host of troubling implications, especially for the members of society who are socioeconomically the worst off<sup>49</sup>. Despite this, a move that often accompanies the decriminalization of sex work in liberal nations is the claim that the reason why sex work should be decriminalized is the claim that “sex work is work” like any other form of service labor, despite the striking and profound degree of harm that affects sex workers as a result of their labor. In opposition to this claim, I grapple in this paper with the question of whether consent to sex work on the part of the sex worker is possible. I first defend the method of analyzing sex work based on it is consensual (as opposed to the existing literature on sex work, which largely assess sex work by whether it is intrinsically harmful in other ways). I offer a set of criteria which any theory of sexual consent must meet, and explain why the existing frameworks of consent used in the debate over sex work do not meet these criteria. Finally, I ultimately propose enthusiasm as a necessary criterion for sexual consent, and then use this criterion to explain why sex work is nonconsensual.

## II. MOVING FROM AN ANALYSIS OF HARM TO AN ANALYSIS OF CONSENT

Much of the existing philosophical literature on sex work revolves around whether the sex work industry is intrinsically or contingently harmful. The argument that sex work is intrinsically harmful to sex workers is often made by feminist thinkers and typically holds that sex work is part of cyclic gendered and economic violence that entraps sex workers and subjects them to enormous physiological harm<sup>50</sup>. The view, often held by contractarian thinkers, that sex work is contingently harmful agrees that sex workers experience tremendous suffering, but argues that this harm is caused by contingent features that are external to sex work, such as the exploitative nature of all wage labor under capitalism and societal stigmas against sex workers<sup>51</sup>. Another hallmark of the existing literature on sex work appeals to the harm done to sex workers—in particular, the material, often consequentialist harms of the profession. For example, authors on both sides of the debate cite the risk to sex workers of assault and battery by clients and pimps, psychological harm like dissociation and depression, and gender

---

<sup>46</sup> Kat Banyard, “The Dangers of Rebranding Prostitution as ‘Sex Work,’” *The Guardian*, June 7, 2016, <https://www.theguardian.com/lifeandstyle/2016/jun/06/prostitution-sex-work-pimp-state-kat-banyard-decriminalisation>.

<sup>47</sup> Banyard, “Dangers.”

<sup>48</sup> Banyard, “Dangers.”

<sup>49</sup> Anderson, “Prostitution and Sexual Autonomy,” 762-66.

<sup>50</sup> Martin Alcoff (2009) describes the feminist view on pg. 127 of “Discourses of Sexual Violence in a Global Framework,” and examples of this view are Dworkin (1993), Westin (2014), McDougall (2014), and Allan (2004).

<sup>51</sup> Examples of authors who hold the contractarian view described here include Ericsson (1980) and Moen (2014).

inequality broadly. I find the debate between the feminist and contractarian views to be thorough and broadly conducted so far, but it lacks definitive resolution from either side.

In contrast to the existing discussion about sex work, I propose adjudicating sex work based on whether or not it is consensual. Sexual consent and its ability to define the boundaries of acceptable sex is a topic that has grown in importance in recent years, especially given the rise in publicity of sexual assault on college campuses and across gendered and economic power imbalances (such as in the *#metoo* movement). Because sex is a constitutive feature of sex work and consent plays a determining role in determining what sorts of sex are permissible, whether or not sex work is consensual also has implications for whether or not sex work is morally permissible. Second, using the framework of consent to adjudicate sex work fits into the existing literature about harm to a large extent. We know from the testimonies of those who have experienced nonconsensual sex that it also causes a great degree of psychological harm, so if sex work is nonconsensual then it is also harmful to the sex worker.

Up to this point, I have defended the project of adjudicating sex work based on whether or not it is consensual (as opposed to whether it is intrinsically or contingently harmful), and I have articulated the need to use a model of sexual consent to adjudicate sex work. I will proceed by identifying the criteria which any model which we use to evaluate sexual consent must meet.

First, our model of consent must reflect the adaptive nature of a person's choice to work as a sex worker. By "adaptive," I mean that our choices can reflect "preference[s] which seems incompatible with the agent's welfare and is causally related to oppressive conditions"<sup>52</sup>. To illustrate an instance in which an agent holds an adaptive preference, Khader gives the example of Bangladeshi women who refuse education so that their sons can receive it, and critiques western feminist scholars (such as Nussbaum and Okin) who hold that the Bangladeshi women make this choice because they lack sufficient agency to make a choice that reflects their true desires<sup>53</sup>. Khader holds that not only does this assessment carry the vestiges of harmful stereotypes about women living in post-colonial societies, it also portrays agents with adaptive preferences as irrational because they are "perpetuating their own oppression without seeming to truly want to"<sup>54</sup>. Similarly to the Bangladeshi women who choose to refuse an education, sex workers' choice to enter the sex work industry is likely incompatible with their welfare (given the multitude of ways in which sex work is physically and psychologically harmful to women) and is often caused by pre-existing, oppressive conditions of poverty and abuse. Acknowledging the adaptive quality of one's choice to become a sex worker is necessary to ensure that we are not overlooking or understating the non-ideal conditions (like child sexual abuse, poverty, substance abuse) that may cause one to become a sex worker. This is necessary for a theory of

---

<sup>52</sup> Serene J. Khader, "Must Theorising about Adaptive Preferences Deny Women's Agency?: Adaptive Preferences and Women's Agency," *Journal of Applied Philosophy* 29, no. 4 (November 2012): 302–17, <https://doi.org/10.1111/j.1468-5930.2012.00575.x>, 310. I extend Khader's understanding of adaptive preferences to an agent's choices here, on the assumption that one's preference for something implies their choosing that thing.

<sup>53</sup> Khader, "Adaptive Preferences," 305.

<sup>54</sup> Khader, "Adaptive Preferences," 302.

<sup>55</sup> Eleanor Milligan and Jennifer Jones, "Rethinking Autonomy and Consent in Healthcare Ethics," in *Bioethics - Medical, Ethical and Legal Perspectives*, ed. Peter A. Clark (InTech, 2016), <https://doi.org/10.5772/65765>, 28. Also, see Linda Martín Alcoff, "Discourses of Sexual Violence in a Global Framework," *Philosophical Topics, University of Arkansas Press* 37, no. 2 (2009): 123–39, <https://doi.org/10.5840/philtopics20093728>, 127.

consent because it acknowledges that people do not make decisions with perfect independence from interpersonal relations and that people make choices with pre-existing attachments in mind<sup>55</sup>.

Second, our model of consent must therefore maintain that sex workers are rational. Khader outlines three features of rationality that a theory of adaptive preferences must: agents must have the *cognitive capacity* to assess the options available to them, may have *complex moral psychology* where they may feel ambivalence about norms which their actions uphold, and reflect that they pick the best possible option out of a *limited option set*. This is necessary, as Khader argues, to ensure that sex workers, or any agents in adaptive contexts, are treated respectfully both in discourse about them and in the real world<sup>56</sup>.

Third, our model of consent must acknowledge the ways in which consenting to sex is both similar and different to consenting to other forms of employment, contracts, and interpersonal exchanges. Intuitively, it seems that our standard for consent should be more robust when the activity in question is sexual, for several reasons. In sexual contexts, participants are deeply emotionally and physically vulnerable in a great number of ways, which explains why rape and sexual violence are emotionally and psychologically traumatic in a more profound way than other instances where a person is coerced or manipulated. As Martín Alcoff puts it, rape “alters subjectivity or self-hood” in a way that more mundane instances of coercion do not<sup>57</sup>. Moreover, decisions about whether or not to consent to individual instances of sex, or what one’s standards are for engaging in sex, are typically highly dependent on the emotional, knee-jerk, and otherwise non-rational inclinations a person has.

In the following sections, I will proceed by assessing the contractarian and feminist models based on our criteria for a framework of consent. I conclude that neither sufficiently meets all criteria, and propose an add-on component of a model of consent that resolves the deficiencies in the feminist model.

### III. THE CONTRACTARIAN MODEL OF CONSENT

As I have mentioned earlier, the contractarian model of sexual consent holds that agents are autonomous in the sense that they can make choices freely and without the influence from others, so any performative agreement that an agent gives to sex reflects a choice to give consent—much like signing a contract. Many of the standard things we hold about consent are a result of the contractarian model’s view of the self as a rational and autonomous actor—for example, that consent must be freely given (not coerced or deceived), and consent must be specific<sup>58</sup>. Importantly, proponents of the contractarian view hold that the necessary conditions for non-sexual consent are not different from the conditions for sexual consent. Applied to sex work, contractarians hold that sex work is harmful, but not intrinsically harmful—only contingently. Arguments for this view generally align with the thrust of argumentation that because sex work is a private economic agreement between consenting adults, it is not intrinsically different from other kinds of service labor. Contractarians hold that the harm to sex workers by their participation in the industry is a result of extrinsic societal

<sup>56</sup> Khader, “Adaptive Preferences,” 305.

<sup>57</sup> Martín Alcoff, “Discourses of Sexual Violence,” 128.

<sup>58</sup> Tom Dougherty, “Sex, Lies, and Consent,” *Ethics* 123, no. 4 (July 2013): 717–44, <https://doi.org/10.1086/670249>, 733.

stigma and criminalization, poor socioeconomic opportunity which condition sex workers into that line of labor, and the general disadvantages to wage labor in a capitalist economy—none of which are intrinsic to sex work<sup>59</sup>.

Upon assessing the contractarian view of consent by the factors outlined in the previous section, we notice several issues. First, the contractarian model of sexual consent is *not adaptive*; it prioritizes the performative utterance of consent and the mental state of intending to consent while overlooking ways in which sex may be either incompatible with the welfare of an agent who consents to it, or causally related to oppressive conditions. Take the instance of a sex and love addict who engages in casual sex: though their consent is legitimate under a contractarian model of consent because they performatively consent and possess the mental state of genuinely desiring casual sex, the model omits the morally relevant information that their desire for sex both sustains and is produced by unhealthy relationship habits. To be clear, the adaptive elements of the sex and love addict's consent does not necessarily void her consent—but, it adds crucial information that can aid our understanding and assessment of nuanced, grey-area cases of consent—such as sex work.

Second, the contractarian model of sexual consent also does not wholly satisfy the rationality desiderata that Khader outlines. It does hold that agents can have the *cognitive capacity* to assess their possible options and choose the best of a *limited option set*, because agents would only cognitively and emotionally consent if they knew that they did not have different or better options. However, the contractarian model does not satisfy the *complex moral psychology* desiderata because conceptualizing of agents who are freely entering and exiting contracts fails to acknowledge that agents' choices are both conditioned by and instrumental in reproducing norms that they can then feel ambivalence about. This ultimately harms the extent to which the contractarian model conceptualizes of agents as rational because it oversimplifies the complex motivations involved in consenting to sex and understates the ways in which agents' sexual choices are contextualized in social norms.

Third, the contractarian model of consent fails to acknowledge the ways in which consenting to sex is different from consenting to other ordinary practices. Instead, contractarians suggest that societies should normalize the commodification of sex so as to destigmatize sex work and mitigate harm to sex workers<sup>60</sup>. I have already defended why accounts of sexual consent need to be more robust than ordinary accounts of consent, but commodified sex in particular would harm agents' autonomy. Should sex be treated like any other service, coercive economic incentives to provide unwanted sex would arise, as would increased government and corporate surveillance and manipulation of peoples' sexual behaviors and desires<sup>61</sup>. This in fact decreases agents' ability to meaningfully consent to sex because it overreaches into their private, subjective determinations of their standards for sexual consent and denies agents control over their own actions<sup>62</sup>.

---

<sup>59</sup> Lars O. Ericsson, "Charges Against Prostitution: An Attempt at a Philosophical Assessment," *Ethics* 90, no. 3 (April 1980): 335–66, <https://doi.org/10.1086/292165>, 362.

<sup>60</sup> Ericsson, "Charges Against Prostitution," 362.

<sup>61</sup> Anderson, "Prostitution and Sexual Autonomy," 763.

<sup>62</sup> Anderson, "Prostitution and Sexual Autonomy," 766.

So, for several reasons the contractarian model of consent is not a sufficient theory of consent by which we can adjudicate whether sex work is consensual. I turn next to the feminist model of consent.

#### IV. THE FEMINIST MODEL OF CONSENT

The feminist model of consent fares slightly better than the contractarian model of consent, but is still deficient in meeting the criteria of explaining what about sex requires consent to it to be especially robust. The feminist model maintains most of the core features of the contractarian view, such as rationality of the agents, non-coercion, non-deception, and specificity of consent, but adds that that an agent can meaningfully consent to an action only if they had other good options which they could have chosen instead<sup>63</sup>. Under this model, if an agent lacks any better options than a particular agreement which they enter, that can also entail a kind of undesirable coercion.

First, the feminist model of consent does meet the *adaptiveness* criteria—it holds that limitations on the set of an agent's options can be caused by oppressive conditions, and that these restrictions on options can box agents into choices that are ultimately detrimental to their own welfare.

Second, the feminist model of consent does meet all three desiderata for rationality. Like the contractarian model, it meets the *cognitive capacity* and *limited option sets* desiderata because it affords that agents can accurately judge their options, determine that their option set is limited, and choose the best of their options. Unlike the contractarian model of consent, the feminist model acknowledges that agents may feel ambivalent about their best choice even though it is their best, and therefore meets the *complex moral psychology* desiderata.

Third, the feminist conception of consent fails to explain why violations of sexual consent are different than the wrongs caused by other forms of wage labor, for example. It is certainly true that limitations on the option sets of sex workers limit the degree to which they freely enter into the sex work industry, but this is true for any worker who takes the best job available to them. Take the case of a coal miner—she may enter the coal mining industry because she has few other jobs available to her, even though coal mining will hugely harm the health of her lungs and the industry has broader negative social effects, such as contributing to air pollution. Like sex work, the coal miner's best option both causes physical harm to her and has problematic social effects—which mirrors the types of harm that the feminist view attributes to sex work. Namely, the feminist view holds that sex work is wrong first because of direct harm to sex workers, and second, because of the broader cyclical structures of disenfranchisement and gendered violence that cause and are caused by sex work. But because coal mining also results in direct harm to the coal miner and an interplay with cyclical structural violence, the feminist model of consent cannot sufficiently explain what is different between the sex worker and the coal miner despite holding that sex work is uniquely disempowering and harmful to the sex worker's autonomy.

---

<sup>63</sup> Anderson, "Prostitution and Sexual Autonomy," 754.

## V. THE ENTHUSIASM CRITERION

Given that neither the contractarian nor the feminist model of consent satisfactorily meets our criteria for a model of consent, I will present a new criterion for sexual consent—the enthusiasm criterion—that, when added to the feminist model of consent, raises the bar for sexual consent and differentiates it from ordinary interpersonal and economic consent. I present the enthusiasm criterion as a necessary condition for the kind of consent that makes sexual activities morally permissible. Then, I will apply the enthusiasm criterion to determine whether sexual consent is really given by sex workers. I conclude that sex workers are intrinsically wronged by their participation in the sex work industry to a greater extent than workers in other industries—such as the coal miner who works in the coal mining industry.

The *enthusiasm criterion* holds that *sexual consent requires an enthusiastic emotional state* on the part of the agent who consents. By this I mean that an agent must have a high level of emotional desire for sex, rather than simply not being of a mind to object to sex. This satisfies our criteria for a model of consent where our model must be *more robust for sexual consent* than for run-of-the-mill instances of consent. In both sexual and non-sexual cases, consent may typically involve positive performative agreement—in other words, we verbally or physically signal that we give our consent to the exchange. Some non-sexual instances of consent where a person performatively consents even though she does not feel enthused about what she has consented to are permissible; for example, I may agree to do chores around my house despite not being enthused about them. But my claim here is that if a person performatively agrees to sex despite not desiring it (or being actively opposed to it), then their performative agreement is not sufficient for morally transformative consent to have occurred. It is much more troubling if a person is not mentally or emotionally enthusiastic in the case of sexual consent than it is for an ordinary instance of consent; in sexual cases, the mental-state element of consent trumps performative agreements we might make.

Then enthusiasm criterion stems from a critique of the view of autonomy which the contractarian position depends upon. The contractarian view holds that subjects are autonomous and takes this to mean that the subject is a “thinking person who is able to make decisions freely, unencumbered by the needs, desires, or perspectives of others”<sup>64</sup>. This legitimizes the contractarian stance that contractual agreements or the giving of performative assent are a sufficient condition for making actions morally permissible. In the context of healthcare ethics, Milligan and Jones push back on this; they claim that it trades off with considerations related to a person’s relationships with themselves and others, falsely equates the bare minimum of treating someone as autonomous with genuine care and respect, and overlooks the fundamentally emotionally vulnerable situation that healthcare patients are often in. In contrast to the view of the subject as rational and capable of making choices that are fully independent, Milligan and Jones hold that humans are “unquestionably and primarily embedded and embodied beings with [...] non-negotiable constraints on their ability to act in a purely individualistic or rationally considered way”<sup>65</sup>. This suggests that we must modify the contractarian view to accommodate for the ways in which non-ideal realities,

<sup>64</sup> Milligan and Jones, “Rethinking Autonomy and Consent,” 22.

<sup>65</sup> Milligan and Jones, “Rethinking Autonomy and Consent,” 24.



like interpersonal and intrapersonal relationships and psychological/emotional vulnerability, often constrain our ability to make fully autonomous choices.

I argue that, when it comes to assessing sexual encounters, this non-ideal conceptualization of autonomy is more appropriate than the fully autonomous contractarian view. Martín Alcoff, describing the phenomenology of sex, writes that “sexuality emerges in intersubjective interaction [...] I know what I want to do fully and with certainty only in the very moment I do it”<sup>66</sup>. This suggests that sexual actions and choices are instinctual, rather than premeditated or chosen through internal deliberation—and therefore are highly dependent on an agent’s emotions, state of mind, and the intersubjective nuances of her relationship or interactions with a sexual partner. Like the non-ideal autonomy framework suggests, sexual choices are not fully autonomous because of how heavily sexual behavior is conditioned by interpersonal and intrapersonal pressures. An agent’s immediate, instinctual reactions in sexual contexts may therefore be more autonomous than their rationalization, which can be influenced by a number of external factors.

In fact, if an agent performatively consents despite being unenthusiastic, it suggests that they might be influenced by manipulation or indirect coercion—particularly in sexual encounters. We can imagine ubiquitous examples where agents performatively agree to sex despite being unenthusiastic because they feel socially obligated—for example, a woman may feel that she “owes sex” to a man who takes her on an expensive date, or people may be peer pressured into seeking out sex that they do not want. In other cases, the interpersonal vulnerability created in sexual scenarios may deter people from backing out of sex. Take for example a person who is not enthusiastic about having sex but realizes that their partner is much more physically powerful than they are, or a person who is not enthusiastic about sex but also wants to avoid hurting their partner’s feelings.

Moreover, sexual situations make agents particularly vulnerable to manipulation and indirect coercion. For one, normative judgements about sex (ie. that one ought, or ought not, wait until marriage) are pervasive and play a large role in the sexual choices that agents make. The same goes for social obligations that people may feel, even if the obligation is not real. Additionally, sexual encounters typically happen in private, which can enable physical coercion and emotional manipulation. Therefore, enthusiasm as a mental state is particularly relevant to sexual consent because it differentiates between autonomous actions and choices and those that are not fully autonomous.

I have so far defended the necessity of enthusiastic consent only in cases of sexual consent because the justifications for it derive from the important features of sexual interactions. Further, I will defend the stronger claim that the enthusiasm criterion does not apply to cases of consent that are not sexual. Returning to the example given in previous sections, I will explain why a lack of enthusiasm on a coal miner’s part does not mean that she does not really consent to working in the industry. The reasons why a coal miner might feel unenthusiastic about her work are largely empirical, in contrast to the various personal emotional, and instinctual reasons which a person might have to be unenthusiastic about sex. For example, coal miners may be rightly unenthusiastic about the prospect of lung disease, grueling work hours, and low pay—in contrast to a person’s intuitions or

---

<sup>66</sup> Martín Alcoff, “Discourses of Sexual Violence,” 128.

instinctual emotional reactions in sexual encounters. This suggests that the reason why work in a coal mine is harmful to the coal miner has less to do with overriding the coal miner's state of mind and has more to do with objectively poor working conditions. In contrast, one's failure to meet the enthusiasm criterion does override the state of mind of the other participant in a sexual encounter, which is particularly pernicious because violations of a person's sexual autonomy entail deep violations of their sense of self, inter- and intra-personal relationships, and emotional experience, as a result of how sexual consent is intimately connected to a person's state of mind. Even though the coal mining industry might be morally impermissible for other reasons—that people have to harm their health and the global climate to make enough money to cover their basic living expenses—working as a coal miner is not *nonconsensual* per the enthusiasm criterion.

Finally, what about cases where a person becomes unenthusiastic during a sexual encounter because it is unenjoyable? In these cases, the agent is not unenthusiastic about the prospect of having sex in general. The issue here is of the specificity of what they consented to, not that they are unenthusiastic. They consented to sex that was enjoyable and what they received was outside the scope of the specific behaviors they consented to<sup>67</sup>. As with any behavior that could plausibly violate sexual consent, this places an onus on agents to communicate with their sexual partner to avoid violations of consent. Should a violation of consent arise in a scenario like this, it is because of a mismatch between the specific behaviors the agent consented to and what actually happened. The enthusiasm criterion is thus not over-limiting, because it does not indict situations like this that are not clear-cut violations of consent.

## VI. APPLYING THE ENTHUSIASM CRITERION

In this section, I address intuitive counterexamples to the enthusiasm criterion and apply the enthusiasm criterion to sex work to explain why it is uniquely harmful. In the previous section, I have compared the example of the coal miner to the sex worker with the aim of explaining why the enthusiasm criterion primarily applies in sexual cases. I offer the coal miner scenario to demonstrate a non-sexual case where a worker is harmed by her work, her work has harmful social effects, and where the worker might strongly prefer a different job if given the opportunity—yet, we likely (and rightly, as I conclude) intuit that this is not fully analogous to the ways in which sex work is harmful.

Before we turn to sex work, I will give an example of applying the enthusiasm criterion to a ubiquitous sexual encounter, in hopes of reiterating important stipulations about the enthusiasm criterion. Take for example a situation where Person A and Person B have gone on a few dates together and like each other quite a bit. Person B wants to have sex with Person A, but Person A feels unsure, nervous, and feels uncomfortable having sex with Person B so soon. Despite this, Person A feigns enthusiasm and gives affirmative consent to Person B. This situation fails to meet the enthusiasm criterion because Person A's mental state was not one of actively desiring sex with Person B. The takeaway here is that enthusiasm is a quality of a mental state, not a quality of the performative agreement or of Person A's behavior.

---

<sup>67</sup> There is more to be said about under what conditions this is a violation of sexual consent, but I will not discuss it here—it is outside the scope of my defense of the enthusiasm criterion.

Imagine also that the day before, Person A's friend has jokingly called them a prude. On previous dates, Person B has remarked that they find sexual compatibility to be a crucial component of overall romantic compatibility. Either consciously or subconsciously, these factors (peer pressure and wanting to remain a viable romantic candidate in Person B's eyes, respectively) may have played a role in Person A's choice to performatively agree to sex despite not being genuinely enthusiastic. The takeaway here is that a lack of enthusiasm can reveal manipulation or indirect coercion when agents agree to sex—in this case, we might say that Person A's giving of performative agreement to sex despite being unenthusiastic was indirectly manipulated or conditioned by a broad range of stigmas and negative connotations surrounding prudishness.

Next, I will address two intuitive counterexamples to the enthusiasm criterion. The first is a case where a person has sex, which they do not enjoy, for the purpose of becoming pregnant. Despite a lack of enthusiasm, this case still meets the enthusiasm criterion due to the specificity of what the wife consents to. While the person is not enthusiastic about casual sex or the physical aspects of sex, she is enthusiastic about conceiving a biological child and sex is a necessary means to that end—therefore, she is sufficiently enthusiastic about sex to pass the enthusiasm criterion. If one is enthusiastic about an outcome, then they must also be enthusiastic about the necessary means to that outcome; here, pregnancy is the outcome which the person who consents to sex is enthusiastic about and sex is a necessary means to that end. The difference between this case and sex work is that even though sex workers may not have other ways to make money besides sex work (and therefore sex work is “necessary” to the end of making money), it is not necessary in the strong physical and biological way that sex is necessary for conceiving a child.

The second case is a case where one person in a long-term relationship has sex with their partner, despite not being enthused about it, because it will make their partner feel loved. On face, this seems similar to the first case because one might think that making their partner feel loved is an outcome of sex with them. However, the difference between this scenario and the first is that making the partner feel loved is not an external outcome of sex but something constitutive to the act of having sex; that is, sex itself would make the partner feel loved. This means that the partner who agrees to sex which they are not enthusiastic about meets the enthusiasm criterion because if sex with their partner and making their partner feel loved are intrinsically connected, then their enthusiasm for making their partner feel loved means that they are also enthusiastic about having sex in the sense relevant to the enthusiasm criterion.

Returning to sex work, I find that in most cases, sex work is not sufficiently consensual because genuine enthusiasm for sex on the part of the sex workers is unlikely. First, women who enter the sex work industry are frequently coerced into it by lack of a better option: “people become and remain sex workers as a result of institutional and familial rejection, and abuse”. Petro goes on to describe the specific conditions that may incentivize one to begin sex work, such as “lack of education [...] entangled in the system, looking for work. Their housing is precarious. Some are in codependent relationships with men they support. Most are women of color. Some are trans.” Absent these abusive conditions that decrease sex workers' economic options and condition them to accept an abusive and physically degrading line of work, sex workers would likely not choose to be sex workers, which demonstrates that authentic enthusiasm on sex workers' part is low.

Second, many studies reveal that female sex workers experience a high frequency of dissociation and other trauma-related psychological disorders, which can cause sex workers to be distanced from their instinctual or autonomous behaviors. As discussed in the previous section, instinctual reactions in sexual scenarios are often the most autonomous behaviors an agent may display. Because the conditions that are risk factors for becoming sex workers, like child sexual abuse, also contribute to dissociation and trauma-related psychological disorders that diminish our ability to differentiate the genuine actions of people from ones that are non-autonomous, it becomes difficult or impossible to maintain that sex workers are enthusiastic about sex work. Even if it were the case that a sex worker felt enthusiastic about physically degrading or psychologically harmful work, we would have great reason to worry that this was a result of psychological conditioning or an unhealthy method of coping with trauma, for example.

Third, sex workers are often entrapped by pimps and are deceived into entering the sex work industry or prevented from leaving. Research from the Urban Institute reports that pimps use manipulative methods “such as flashing money around, seducing them, entering into romantic relationships, convincing them that they may as well make money if they’re already having sex, or having other female employees sell the idea [of becoming a sex worker]”. Around 15 percent of pimps—likely a low estimate due to self-reported data—report using violence against sex workers at some point. Further complicating this issue is the heightened and increasing policing of sex work, despite a number of clients being police officers and even cases when police officers abuse their status to extort sexual and monetary favors—which creates a double bind where sex workers must either endure unsafe working conditions or exchange them for a criminal justice system that might be just as bad or worse. Gaps in sex workers’ employment records, addictions, and having past criminal convictions may prevent sex workers from finding jobs outside of the industry. The ‘happy hooker’ is just a trope; in reality, sex workers enter the industry and remain in it for reasons that are far from enthusiasm.

In this paper, I have defended the importance of focusing on consent (or lack thereof) as a way in which harm is done and experienced. I have then outlined the theories of consent underlying both sides of the debate over sex work, and concluded that neither of them constitutes a complete theory of sexual consent. I have then proposed and defended the enthusiasm criterion as a mental-state condition for sexual consent, and explained why sex work fails to meet this criterion. My final conclusion is that sex work is not fully consensual in a morally relevant way, which should prompt us to reconsider the ethicality of the sex work industry writ large.

## BIBLIOGRAPHY

Alcoff, Linda Martín. "Discourses of Sexual Violence in a Global Framework:" *Philosophical Topics*, University of Arkansas Press 37, no. 2 (2009): 123–39. <https://doi.org/10.5840/philtopics20093728>.

Allan, Charlotte. "The Happy Hooker?" *The British Journal of General Practice* 54, no. 504 (July 1, 2004): 556–57. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1324819/>.

Anderson, Scott A. "Prostitution and Sexual Autonomy: Making Sense of the Prohibition of Prostitution." *Ethics* 112, no. 4 (July 2002): 748–80. <https://doi.org/10.1086/339672>.

Banyard, Kat. "The Dangers of Rebranding Prostitution as 'Sex Work.'" *The Guardian*, June 7, 2016. <https://www.theguardian.com/lifeandstyle/2016/jun/06/prostitution-sex-work-pimp-state-kat-ban-yard-decriminalisation>.

Dougherty, Tom. "Sex, Lies, and Consent." *Ethics* 123, no. 4 (July 2013): 717–44. <https://doi.org/10.1086/670249>.

Dworkin, Andrea. "Prostitution and Male Supremacy." *Michigan Journal of Gender & Law* 1, no. 1 (1993): 13.

Ericsson, Lars O. "Charges Against Prostitution: An Attempt at a Philosophical Assessment." *Ethics* 90, no. 3 (April 1980): 335–66. <https://doi.org/10.1086/292165>.

Johnson, Matthew, and Meredith Dank. "The Hustle: Economics of the Underground Commercial Sex Economy," 2014. <http://apps.urban.org/features/theHustle/index.html>.

Khader, Serene J. "Must Theorising about Adaptive Preferences Deny Women's Agency?: Adaptive Preferences and Women's Agency." *Journal of Applied Philosophy* 29, no. 4 (November 2012): 302–17. <https://doi.org/10.1111/j.1468-5930.2012.00575.x>.

Liberto, Hallie. "Intention and Sexual Consent." *Philosophical Explorations* 20, no. sup2 (August 31, 2017): 127–41. <https://doi.org/10.1080/13869795.2017.1356359>.

Marino, Patricia. "Philosophy of Sex: Philosophy of Sex." *Philosophy Compass* 9, no. 1 (January 2014): 22–32. <https://doi.org/10.1111/phc3.12098>.

McDougall, Rosalind J. "Intrinsic versus Contingent Claims about the Harmfulness of Prostitution." *Journal of Medical Ethics* 40, no. 2 (February 2014): 83–83. <https://doi.org/10.1136/medethics-2012-100833>.

Milligan, Eleanor, and Jennifer Jones. "Rethinking Autonomy and Consent in Healthcare Ethics." In *Bioethics - Medical, Ethical and Legal Perspectives*, edited by Peter A. Clark. InTech, 2016. <https://doi.org/10.5772/65765>.

Moen, Ole Martin. "Is Prostitution Harmful?" *Journal of Medical Ethics* 40, no. 2 (February 2014): 73–81. <https://doi.org/10.1136/medethics-2011-100367>.

———. "Prostitution and Sexual Ethics: A Reply to Westin." *Journal of Medical Ethics* 40, no.2 (February 2014): 88–88. <https://doi.org/10.1136/medethics-2013-101902>.

North, Anna. "Sex Workers' Fight for Decriminalization, Explained." Vox, August 2, 2019. <https://www.vox.com/2019/8/2/20692327/sex-work-decriminalization-prostitution-new-york-dc>.

Ozymandias. "Against 'Enthusiastic Consent.'" *Thing of Things* (blog), July 7, 2016. <https://thingofthings.wordpress.com/2016/07/07/against-enthusiastic-consent/>.

Petro, Melissa. "Getting Out of Sex Work." Pacific Standard. Accessed August 20, 2020. <https://psmag.com/economics/getting-out-of-sex-work>.

Tschoeke, Stefan, Raoul Borbé, Tilman Steinert, and Dana Bichescu-Burian. "A Systematic Review of Dissociation in Female Sex Workers." *Journal of Trauma & Dissociation* 20, no. 2 (March 15, 2019): 242–57. <https://doi.org/10.1080/15299732.2019.1572044>.

Wertheimer, Alan. "Consent and Sexual Relations," n.d., 25.

Westin, Anna. "The Harms of Prostitution: Critiquing Moen's Argument of No-Harm." *Journal of Medical Ethics* 40, no. 2 (February 2014): 86–87. <https://doi.org/10.1136/medethics-2012-101082>.

---

---

What to Make of  
Medical Nihilism?  
An Epistemological  
Defense of Optimism

---

---

*Stanley Bowen Zhang*  
*University of Toronto*

## FOREWORD

*Editors Ashley Gasdow and Leonardo Villa-Forte*

In ‘What to Make of *Medical Nihilism*? An Epistemological Defense of Optimism’, Stanley Bowen Zhang responds to this compelling and relevant question by providing a defense of trust in modern medicine. Zhang begins by outlining an argument against having trust in medicine, set forth by Jacob Stegenga. Despite Stegenga’s argument, Zhang says, we have strong epistemic reasons for retaining our trust in medicine. To this end, Zhang touches on a variety of key ideas such as the role of procedural clinical knowledge on medical practice, and how physician metaknowledge contributes to systems that advise them on diagnoses and treatments. After identifying kinds of knowledge that don’t fall under Stegenga’s argument and describing how improvements in medical metaknowledge contribute to technologies that assist professionals, Zhang argues that Stegenga’s argument is itself informed by kinds of medical metaknowledge.

We selected this paper because it presents a cogent, persuasive argument that uses interesting and relevant examples, making for a compelling discussion of these topics. There is no doubt that our contemporary landscape, with the contextually unique challenges it has posed, has illuminated modern medicine now more than ever before as a central fixture. Living through a pandemic has placed doubt in the minds of many as to the efficacy of medicine and the scope of its present-day knowledge. Yet, on the other hand, we have also seen unprecedented scientific developments, ones which will undoubtedly have far-reaching implications and a great impact on our lives.

## I. INTRODUCTION

In *Medical Nihilism*, Jacob Stegenga makes a compelling case against having a trusting attitude towards modern medicine. His Bayesian “master argument” encourages a very low degree of confidence in the effectiveness of any novel medical intervention, even when there is available scientific evidence in favour of the proclaimed efficacy. According to Stegenga, the “central ambition” of his work is to contribute to current debates on contextualized demarcation in the philosophy of medicine; he does this by exposing the underlying biases and malleability of medical research methods, the theoretical shortcomings of evidence-based medicine (EBM), and problems with the standards of clinical practice prescribed by EBM.<sup>68</sup> However, Stegenga would agree that the rational response to most serious medical concerns—be it an acute injury or a chronic health problem—should still be to consult the advice of medical experts. Although the final chapter<sup>69</sup> of *Medical Nihilism* offers reasonable proposals for how we can realign medical research in ways consistent with Stegenga’s skeptical conclusions, the book itself does not seem to provide enough contextualized clarifications about what the “nihilist” thesis really says about medical knowledge in general. Thus, many readers may interpret Stegenga’s thesis to mean that we should not

---

<sup>68</sup> Jacob Stegenga, *Medical Nihilism* (Oxford: Oxford University Press, 2018), 1–18.

<sup>69</sup> Stegenga, *Medical Nihilism*, 185–97.



have confidence in the legitimacy of medical knowledge altogether. The use of the term “nihilism” exacerbates this worry.

In this essay, I argue that even if Stegenga’s arguments succeed, there remain strong epistemological reasons to maintain confidence in—and an optimistic attitude towards—the legitimacy and reliability of medical knowledge in general. To this end, I will elucidate the ways in which professional medical consultation and practice are informed by several different types of medical knowledge, and I will argue that Stegenga’s arguments only call into doubt a restricted subtype of *propositional* medical knowledge. In particular, procedural clinical knowledge and medical metaknowledge almost entirely escape Stegenga’s “nihilism.” Reliable qualitative medical metaknowledge (which will be explained in much detail) also leads to more reliable professional judgement, better medical technologies, and higher-quality medical consultation. Finally, I will argue that the Bayesian “master argument” in *Medical Nihilism*—and, in general, skeptical arguments that are structurally similar to it—must in fact rely upon a significant expanse of qualitative, empirically grounded medical metaknowledge, and that these skeptical arguments, if successful, contribute positively to medical metaknowledge.

## II. ON PROCEDURAL CLINICAL KNOWLEDGE

Early epistemologists used to define knowledge in terms of what it is to know the *truth* of a given belief, which is always expressible as a proposition. What this classical conception of knowledge describes is also called *propositional* or *constative* knowledge.<sup>70</sup> While a pre-modern outlook might hold that this concept accounts for all knowledge, most philosophers today would instead agree with a distinction between “knowledge-that” and “knowledge-how” often credited to Gilbert Ryle: the latter category, also known as *procedural* or *imperative* knowledge, refers to knowledge that cannot be reduced to declarative propositions.<sup>71</sup>

For example, when a patient enters a clinic for a blood test, the phlebotomist will know the section of the patient’s upper arm around which a tourniquet needs to be tightened; she will also know how to clean the site of the injection to reduce the risk of infection, as well as how to locate a vein for the insertion of the needle. Similarly, when a clinical pathologist is taking a bacterial culture, she knows how to differentiate various cellular structures when looking into a microscope, what certain species of bacteria look like, and when to perform relevant diagnostic tests based on what she sees. Even if we can completely describe what “knowledge” facilitates these competences in terms of clearly stated instructions, these corresponding statements will always be *prescriptive* rather than *descriptive* (i.e., propositional).<sup>72</sup> From the perspectives of those who possess the learned capacity to reliably perform these procedures, the *content* of their knowledge is not exhausted by the descriptions of the required actions. In other words, there must be a fundamental difference between knowing a set of instructions and knowing how to carry out that set of instructions, especially in such complex professional

<sup>70</sup> R. E. Ashcroft, “Current Epistemological Problems in Evidence Based Medicine,” *Journal of Medical Ethics*, no. 30 (April 2004): 131–32.

<sup>71</sup> Ashcroft, “Current Epistemological Problems,” 132.

<sup>72</sup> Kazem Sadegh-Zadeh, “Types of Medical Knowledge,” in *Handbook of Analytic Philosophy of Medicine*, (Dordrecht: Springer Science, 2015), 471.

settings. If this were not the case, then practicums and laboratory experience would not have been an integral component of the formal education of medical scientists. Some epistemic content involved in the latter kind of “knowing” cannot be reduced to propositional knowledge. These non-propositional components are what modern epistemologists refer to as *procedural* or *imperative* knowledge.

What *Medical Nihilism* sets out to do for the most part is to question the general plausibility of assertions claiming efficacy for novel interventions by relying on a given expanse of second-order evidence—that is, evidence about the nature of the first-order evidence used to vindicate the sorts of medical-scientific propositions that *Medical Nihilism* calls into doubt—for the presence of malleability and bias in the methods and processes by which said medical intervention is evaluated for approval purposes. There are many noteworthy epistemological consequences of this realization. For now, we should note that this approach only amounts to doubting a particular type of first-order propositional medical knowledge at best.

EBM, on the other hand, can be seen as a theoretical account of what purportedly *should be* considered as legitimate and reliable propositional medical knowledge. Despite the significant authoritative impact that EBM has had on medical practice, its framework mostly accounts for understandings of supposedly dependable standards of clinical practices that can be reduced to propositional knowledge.<sup>73</sup> It should be noted that Stegenga’s thesis should not be interpreted as antithetical to the foundations of EBM. The brunt of Stegenga’s skeptical push is directed towards the cases where medical interventions are not truly warranted by rigorous science, with regard to the ideal of scientific rigour that underlie EBM. However, these cases abound also arguably because of shortcomings in the formulation of EBM’s standards—or, for example, the community-endorsed but under-examined trust in the evidential significance of randomized controlled trials and the meta-analyses of those trials.<sup>74</sup> The kinds of medical knowledge for which EBM can serve as an epistemic foundation or framework thus cover the overwhelming majority of the specific sorts of assertions that Stegenga calls into doubt in *Medical Nihilism*. This also means that Stegenga’s “nihilism” does not target the aforementioned kind of non-propositional, *procedural* medical knowledge.

Throughout their educational and professional careers, physicians and medical scientists amass significant procedural clinical knowledge. Procedural knowledge abounds in practical medicine and constitutes the foundation upon which medical practice is grounded, shaping how medicine is learned and taught in ways arguably more profound than propositional medical knowledge.<sup>75</sup> Ultimately, this speaks to the fact that medicine is a practical science, rather than a natural or formal science.

Perhaps some would argue that *Medical Nihilism* might also *implicitly* target some procedural clinical knowledge, for the effectiveness of some clinical procedures would seem to be susceptible to the same sort of skeptical challenges as the effectiveness of—for instance—pharmaceutical interventions. However, even if this were true, such cases are not genuine examples of the type of procedural knowledge I am discussing here; the claim that a given procedure achieves a certain goal is still a propositional claim. Instead, I am referring to

---

<sup>73</sup> Ashcroft, “Current epistemological Problems,” 132.

<sup>74</sup> Stegenga, *Medical Nihilism*, 5–15.

<sup>75</sup> Sadegh-Zadeh, *Handbook*, 472.

non-propositional *epistemic content* involved in the ability to carry out complex procedures believed to serve important functions, not constative beliefs about whether a given procedure is effective. The kind of procedural biomedical knowledge that escapes Stegenga's "nihilism" is an important and foundational part of medical knowledge, and it is therefore also a significant (but perhaps indirect) source of our confidence in the legitimacy of medicine, as a discipline capable of producing genuine scientific knowledge.

### III. ON MEDICAL METAKNOWLEDGE

Other than basic scientific facts and the sort of procedural clinical knowledge described hereinbefore, the knowledge of an experienced medical scientist or physician also include, for example, the merits and shortcomings of accepted rules of practice—as consequences of the virtues and inadequacies of the theoretical frameworks behind these standards; the different limitations on—and variability of—their own judgements and capacities for domain-specific pattern recognition; and—for many experts—the distinctions between different types of medical knowledge. These are (or induce by inference) examples of propositional knowledge about medical knowledge—or the nature of what is believed to be medical knowledge. Such knowledge is often referred to as medical *metaknowledge*, which also escapes Stegenga's "nihilism" for the most part. I further qualify and elaborate on this point in the next section. For now, it is important to note that much of the philosophy of medicine is done in pursuit of medical metaknowledge.<sup>76</sup>

In medical practice and consultation, metaknowledge about how different types of scientific knowledge come together often (implicitly) plays a pivotal role. It should be clear that, for example, metaknowledge about the ways in which certain mechanistic knowledge pertaining to the physiological intricacies of different demographic groups can lead to better interpretations of available empirical evidence for how a disease responds to different treatments, where the empirical evidence consists of otherwise purely propositional first-order knowledge about specific samples. My lingering worry here is that these rather simple (or perhaps minimal) accounts and examples (of the importance and helpfulness of metaknowledge) may *not* be sufficiently robust and compelling to mitigate the momentous reduction in confidence and optimism that *Medical Nihilism* is likely to instigate among its readers. In this section, I will explain some precise ways in which improvements in reliable qualitative medical metaknowledge also lead to better medical technologies that improve the overall quality of medical consultation and diagnostics. The goal here is to show that medical metaknowledge achieves something very concrete in medical science, not just vague improvements in the self-awareness of medical experts.

Today, explicit statements of medical metaknowledge are used to define the structure and logic that underlie the architecture of artificial knowledge-based clinical decision support systems, including so-called "medical expert systems," which are designed to give problem-specific advice to physicians and scientists.<sup>77</sup> Examples of these expert systems include MYCIN, INTERNIST-1, QMR (Quick

<sup>76</sup> Sadegh-Zadeh, *Handbook*, 478.

<sup>77</sup> Sadegh-Zadeh, *Handbook*, 478.

Medical Reference), and DXplain. MYCIN, for instance, not only helped physicians select the appropriate antimicrobial therapy for patients with meningitis and blood infections but also suggested diagnoses, offered explanations, and even requested additional data when the data available to the system was deemed inadequate; it outperformed junior doctors and competed with domain experts.<sup>78</sup> DXplain, on the other hand, is based on a modified form of Bayesian logic and has a database that stores precise descriptions of over 2400 diseases associated with over 5000 clinical manifestations (i.e., signs, symptoms, and findings). Early studies of the reliability of its diagnoses in 105 cases found that it agreed with *post hoc* expert diagnoses in 91% of the cases.<sup>79</sup>

The functionality of these systems primarily rely on a domain-specific “knowledge base”<sup>80</sup> and an “inference engine”, the latter of which is a control structure that determines how the system will approach formal problem-solving.<sup>81</sup> The content of any such a knowledge base typically consists of data-type representations of propositional medical knowledge drawn from scientific literature. For these systems to make use of their knowledge bases, “knowledge engineering” is a crucial task. This involves eliciting relevant domain-specific knowledge from literature and representing said knowledge as statements in special-purpose formal languages amenable to “computational reasoning”—or algorithmically executable syntactic-semantic transformations.<sup>82</sup>

Medical metaknowledge, specifically metaknowledge about how distinct kinds of first-order knowledge of different domains of expertise often come together to inform medical practice, plays an important role in this process; it allows these expert systems to convert data into “judgements” such as diagnoses, explanations, and recommendations of viable treatments or further tests. For example, suppose the knowledge engineer knows how to store within a knowledge base the propositional knowledge from literature on cardiology that coronary heart disease is frequently accompanied by angina pectoris in adult patients; without the clinical experience of an expert physician, the computer program requires a more restricted quantification of what is meant by the vague adverb “frequently.” The knowledge engineer may want to specify a set of rules according to which the system should consult the knowledge base for data obtained from empirical studies of the correlation between angina pectoris and coronary heart disease. This would involve several kinds of medical metaknowledge. On the one hand, the assembly of the domain-specific knowledge base requires some knowledge about the design or methodological underpinnings of the studies that divulged the given statistical correlation, as this would factor significantly into how the empirical results are to be interpreted and represented in the knowledge base.<sup>83</sup> On the other hand, knowledge about the structure of professional language in medicine, including ambiguities and subtleties around specific terminologies, is vital not only in the processing of medical literature for the knowledge base, but also in the translation of key findings into a purpose-specific formal language.<sup>84</sup>

---

<sup>78</sup> Kazem Sadegh-Zadeh, “Clinical Decision Support Systems,” in *Handbook of Analytic Philosophy of Medicine*, (Dordrecht: Springer Science, 2015), 717–18.

<sup>79</sup> Sadegh-Zadeh, *Handbook*, 718–19.

<sup>80</sup> Sadegh-Zadeh, *Handbook*, 719.

<sup>81</sup> Sadegh-Zadeh, *Handbook*, 721.

<sup>82</sup> Sadegh-Zadeh, *Handbook*, 720.

<sup>83</sup> Sadegh-Zadeh, *Handbook*, 719–20.

<sup>84</sup> Kazem Sadegh-Zadeh, “Medical Linguistics,” in *Handbook of Analytic Philosophy of Medicine*, (Dordrecht: Springer Science, 2015), 61–63.

All these dissimilar aspects of the elicitation and representation of first-order knowledge cannot be combined into a coherent consultation system without a sophisticated decision-making and problem-solving framework. Such frameworks dictate how these consultation systems “reason.” The foregoing discussions are meant to illustrate how the architecture of such a framework relies on metaknowledge about first-order medical knowledge that, in most cases, can only be provided by domain experts. As mentioned earlier, each expert system has a control structure, called the inference engine, which is essentially a programmatic implementation of such a framework. Medical metaknowledge thus plays a pivotal role in the architecture of the inference engine, which commits to a particular system of logic depending on the problem domain for which the expert system is built.

To see this point more clearly, it may be helpful to look at how the inference engine uses various kinds of formal logic studied in philosophy. For example, it may rely on temporal logic if the union between patient data and data in the knowledge base deals with time periods and series; it may use deontic logic if that union set contains information expressed in deontic sentences; it may also use alethic modal logic if it needs to deal extensively with possibility and necessity.<sup>85</sup> All these kinds of formal logic help specify *procedural instructions* for the inference engine: For example, given the a set of patient data, if we want to estimate the probability that the patient suffers from condition X, then you ought to conduct tests  $Y_1, Y_2, \dots, Y_n$  and look for whether the outcomes of these tests are  $Z_1, Z_2, \dots, Z_n$ , respectively. Mirroring what I discussed much earlier, this also speaks to the fact that even first-order clinical knowledge is usually procedural, or knowledge-how, rather than propositional, especially in diagnostics.<sup>86</sup> The reliability of these machine-generated diagnostics come from reliable metaknowledge about what procedural knowledge is the most applicable to a given case, taking into account the limitations and advantages of the specific set of propositional knowledge in the knowledge base.

Finally, all kinds of medical metaknowledge hereinbefore described are impervious to Stegenga’s “nihilism,” and we saw compelling reasons to remain confident in the benefits they bring to medicine. They enhance the overall reliability of professional judgement, medical technologies, and medical consultation. In the next part, I will explain how and why Stegenga’s *Medical Nihilism* in fact contributes positively to useful medical metaknowledge of this sort.

#### IV. THE “NIHILIST’S” CONTRIBUTIONS TO METAKNOWLEDGE.

I want to first qualify my position here and concede that medical metaknowledge may not escape Stegenga’s “nihilism” in its entirety, because “metaknowledge” can technically refer to any knowledge about any pre-selected knowledge. Thus, this concept inevitably harbours some degree of equivocality. For example, in medical research, “meta-analyses” refer to formal quantitative studies designed to systematically assess prior studies, usually randomized controlled trials. These are consolidated reviews of possibly conflicting bodies of literature. Much of the conclusions derived from them can also be considered

<sup>85</sup> Kazem Sadegh-Zadeh, “The Logic of Medicine,” in *Handbook of Analytic Philosophy of Medicine*, (Dordrecht: Springer Science, 2015), 702–703.

<sup>86</sup> Sadegh-Zadeh, *Handbook*, 704.

metaknowledge, and they usually lead to quantitative ratings of the consistency or reliability of inferences from clinical research.<sup>87</sup> In Chapter 6 of his book, Stegenga exposes the malleability of these meta-analyses and argues that a wide range of subjective decisions can be made to significantly bias the results.<sup>88</sup> Therefore, this is one type of metaknowledge that does not escape Stegenga's "nihilism."

However, the primary goal of meta-analyses is to obtain estimated measures of the effect of a treatment, and conclusions from these studies are mainly quantitative, which does not fit the same idea of *qualitative* or *epistemological* metaknowledge described in my earlier arguments. Stegenga's contention is primarily that the *quantitative* aspects of the results from these meta-analyses should not be blindly trusted when used as comparative tools. Nevertheless, what surely escapes Stegenga's "nihilism" must include the kind of metaknowledge that informs his own arguments, such as knowledge about the methodological malleability and systematic bias in these meta-analyses (which is, of course, qualitative knowledge about the nature of the knowledge—or what is believed to be knowledge—gained from the results of meta-analyses).

Here, I argue that if the main arguments in *Medical Nihilism* succeed, then Stegenga makes a positive contribution to this *qualitative* kind of metaknowledge. Roughly speaking, this is simply because (1) the premises for Stegenga's Bayesian "master argument" are largely based on this sort of medical metaknowledge and (2) the argument is deductive and valid. The validity simply follows from Bayes' theorem in probability theory. My point is just that if it is sound, then it actually contributes to one aspect of medical knowledge while seemingly undermining another (far narrower) aspect of it. In particular, even if it successfully establishes reasons for us to lower our confidence in the efficacy of *novel* medical interventions, it nevertheless helps the sort of medical metaknowledge that we know is highly useful in enriching professional judgement and enhancing medical consultation technologies.

To see this point more clearly, let us recapitulate how the "master argument" was formalized and how its premises are supported. Let  $H$  be the hypothesis that a given novel medical intervention is efficacious. Then  $\mathbf{P}(H | E)$  is the conditional probability that  $H$  holds true given the evidence  $E$  in support of  $H$ . Bayes' theorem tells us that  $\mathbf{P}(H | E)$  is equal to  $\mathbf{P}(E | H) \times \mathbf{P}(H) / \mathbf{P}(E)$ , where  $\mathbf{P}(E | H)$  is the conditional probability of observing  $E$  if  $H$  is in fact true;  $\mathbf{P}(H)$  is the prior probability of  $H$ , or our prior confidence in  $H$ ; and  $\mathbf{P}(E)$  is the prior probability of observing  $E$ .<sup>89</sup> What I mean by the "main ideas" of Medical Nihilism are thus reasons to think that the value of the denominator  $\mathbf{P}(E)$  is high while both  $\mathbf{P}(E | H)$  and  $\mathbf{P}(H)$ , in the numerator, are low; so that our confidence in  $H$ , even given the evidence  $E$ , ought to be low.

For a significant portion of the book, Stegenga argues that  $\mathbf{P}(E)$  is high due to the malleable nature of biomedical research methods. In other words, biomedical research methods are, in some sense, designed to make interventions appear safer and more effective than they truly are. At various stages of the process of confirming a hypothesis like  $H$ , the observable results are usually skewed towards observing data in favour of  $H$  because of the ubiquity of biased measuring instruments,

<sup>87</sup> A. B. Haidich, "Meta-analysis in Medical Research," *Hippokratia*, 14 (2010, Suppl 1): 29–37.

<sup>88</sup> Stegenga, *Medical Nihilism*, 84–97.

<sup>89</sup> Stegenga, 18.

biased subject recruitment practices, data dredging, and publication bias.<sup>90</sup> This sort of confirmation bias is often attributable to the enormous financial incentive in pharmaceutical research funded by for-profit pharmaceutical companies.<sup>91</sup>

On the other hand, our prior confidence in H before observing the evidence E, denoted  $\mathbf{P}(H)$ , should be low because of the historical paucity of truly successful interventions<sup>92</sup> and the multifaceted pathophysiological complexity behind the casual bases of diseases that renders truly successful interventions highly improbable.<sup>93</sup> This is corroborated by empirical evidence about how small a percentage of experimental interventions really succeeded in history, and this also echoes the well-known fact that the most serious adverse effects of intrusive interventions are often unforeseen.

Finally,  $\mathbf{P}(E | H)$  is low because of typically diminutive effect sizes and discordant evidence. Recall that  $\mathbf{P}(E | H)$  represents the probability of seeing the kind of evidence that we actually observe *if H were true*. In the most recent three to four decades, a large number of therapeutic interventions under investigation have diminutive effect sizes, which means that it is difficult to get these interventions to appear more effective than placebos in the most carefully conducted randomized controlled trials; to make the case for this, Stegenga examines a large number of examples of relatively common interventions, such as antidepressants, tamoxifen, methylphenidate, statins, alendronate, and oseltamivir.<sup>94</sup> We would not expect this sort of empirical findings *if these interventions were truly effective*. In other words, if H were true (in our generic hypothetical case), then the probability of seeing the sort of evidence in support of H that we typically *do* observe would be low. Furthermore, if the total possible space of E includes some consistent evidence  $E_c$  supporting H *as well as* evidence *against* H, and if the biases in medical research are such that it is, on average, more likely to observe what belongs in  $E_c$  than it is to observe what belongs in E, then in most cases we would have  $\mathbf{P}(E | H) < \mathbf{P}(E_c | H)$ , and we should think that  $\mathbf{P}(E_c | H)$  is already very small.<sup>95</sup>

The facts and reasons behind these aforementioned premises are all knowledge about the nature of particular sorts of propositional beliefs held by professionals of the medical sciences. Believing in the soundness of this “master argument” would amount to believing that we have genuine knowledge *about the nature of* certain aspects of medical knowledge—in particular, about the malleability of medical research methods by which we both put forth and confirm our hypotheses, the ubiquity of systematic bias in early stages of this process, the epistemic difficulties involved in uncovering accurate causal relations from complex physiological processes that underlie most medical conditions in need of treatment, and the profusion of discordant evidence that we may selectively overlook when making judgements. Genuinely having knowledge of any of these things would amount to having true, coherent, and reliable medical metaknowledge of the qualitative kind specified earlier.

---

<sup>90</sup> Stegenga, *Medical Nihilism*, 71–149.

<sup>91</sup> Stegenga, *Medical Nihilism*, 162.

<sup>92</sup> Stegenga, *Medical Nihilism*, 171–74.

<sup>93</sup> Stegenga, *Medical Nihilism*, 65–7.

<sup>94</sup> Stegenga, *Medical Nihilism*, 167–75.

<sup>95</sup> Stegenga, *Medical Nihilism*, 176–78.

Therefore, if we accept the Bayes' theorem and the legitimacy of Stegenga's application of it, and if we grant that this *deductive* "master argument" is sound, then we must also accept the underlying metaknowledge that sustains these premises. This means that the conclusion at which we arrive should thus be considered a legitimate, true, consistent, and thus *positive* addition to medical metaknowledge. On this account, most of medical metaknowledge is not only impervious to Stegenga's "nihilism," but also enriched by his in-depth review of the sources of epistemic inadequacies in medical research. More importantly, the "master argument" itself presupposes an expanse of empirically grounded medical metaknowledge.

## V. CONCLUSION

We saw that the judgements and recommendations of physicians and medical scientists are informed by propositional medical knowledge, procedural medical knowledge, and medical metaknowledge, all of which contribute to the legitimacy and reliability of medical knowledge in fundamental ways. More importantly, the arguments in this paper demonstrate that, with respect to the dependability of medicine and medical knowledge, skeptical arguments of the sorts found in *Medical Nihilism* do not apply to, but in fact presuppose, a certain expanse of expert knowledge, or more specifically, empirically grounded medical metaknowledge. To some extent, the better-supported and the more compelling the skeptical argument is, the broader and deeper this presupposition of knowledge tends to be. As discussed much earlier, these arguments do not apply to procedural clinical knowledge, either. Therefore, from an epistemological perspective, it is still highly rational to remain confident in the enormous practical benefits that qualitative medical meta-knowledge can bring to clinical practice, medical consultation, and medical technologies. Furthermore, as a work in the philosophy of medicine, *Medical Nihilism* contributes positively to this kind of medical metaknowledge, if it is a truly successful work. Therefore, even if we really should have low confidence in the effectiveness of novel medical interventions because of the arguments found in *Medical Nihilism*, the rational epistemic attitude towards medicine in general should still be substantially more optimistic than what the term "nihilism" may suggest.



BIBLIOGRAPHY

- Ashcroft, R. E. "Current Epistemological Problems in Evidence Based Medicine." *Journal of Medical Ethics*, no. 30 (April 2004): 131–135.
- Haidich, A. B., "Meta-analysis in Medical Research," *Hippokratia*, 14 (2010, Suppl 1): 29–37.
- Sadegh-Zadeh, Kazem. "Clinical Decision Support Systems." In *Handbook of Analytic Philosophy of Medicine*, 715–732. Dordrecht: Springer Science, 2015.
- Sadegh-Zadeh, Kazem. "Medical Linguistics." In *Handbook of Analytic Philosophy of Medicine*, 61–68. Dordrecht: Springer Science, 2015.
- Sadegh-Zadeh, Kazem. "The Logic of Medicine." In *Handbook of Analytic Philosophy of Medicine*, 699–705. Dordrecht: Springer Science, 2015.
- Sadegh-Zadeh, Kazem. "Types of Medical Knowledge." In *Handbook of Analytic Philosophy of Medicine*, 463–478. Dordrecht: Springer Science, 2015.
- Stegenga, Jacob. *Medical Nihilism*. Oxford: Oxford University Press, 2018.



---

---

# Is Western Philosophy Irredeemably Racist?

---

---

*Shariq Haidery*  
*St. Hugh's College, University of Oxford*

FOREWORD

*Editor Meghana Gavirneni*

In “Is Western philosophy irredeemably racist?” Shariq Haidery answers this question with yes, Western philosophy is racist, but not so in a way that is irredeemable. Haidery defends their position by providing various arguments for both of these points and concludes their paper by proposing a solution for how western philosophy can overcome its racism. Haidery suggests that a global approach should be adopted that integrates historical context while examining philosophers and their ideas. Finally, Haidery explains why this paper can be considered philosophy.

Ultimately, this paper stood out among the rest because of its originality. Haidery takes a meaningful, complex, and highly relevant question and provides the reader with an insightful answer that leaves us further pondering the question ourselves. Thereby, Haidery succeeds in writing an engaging and clear work that will captivate any reader, even one unfamiliar with philosophy. Furthermore, we applaud Haidery for the ambitious argument they produce and their ability to bravely tackle this topic.

By encouraging a critical examination into Western philosophy, this paper asks all of us to consider the biases that may be impacting the world around us and encourages us to truly question all that we are learning.

ABSTRACT

My answer to the titular question is in short no. This is because of mainly the inquisitive nature of philosophy, which means western philosophy can challenge its ahistoricity. Thus, western philosophy can use historical context to understand how a philosopher’s racism may have informed their work and influenced the world. This essay has four parts. Firstly, there is a discussion about the reasons why western philosophy is racist. Those reasons, namely being, the ahistorical aspect of western philosophy, which means that western philosophy cannot confront the racism of its most prominent thinkers; the racism behind much of western political philosophy used to subjugate non-white people. Secondly, I outline the three reasons why philosophy cannot be considered irredeemably racist. Those being: the inquisitive nature of philosophy; the fact western philosophy has been used as a framework to develop liberating ideas for marginalised groups alongside those used to subjugate them; some areas of western philosophy are indeed separate from the issue of race and racism. Thirdly, I turn my attention to how western philosophy may overcome its racism through a global approach that integrates historical context when examining philosophers and their ideas, to prevent the encroachment of discriminatory biases. Finally, this essay responds to an important counter-argument of the essay: how is this essay philosophy?

I. INTRODUCTION

This essay will argue that western philosophy is indeed racist, but that it is not irredeemably racist. It will be split into four parts. Firstly, two reasons why western philosophy is racist will be discussed. Those two reasons being: the

ahistorical nature of western philosophy that has prevented it from engaging in a critical examination of racist philosophers and their influential and pervasive racist ideas; the fact that parts of philosophy (for example social contract theory) in Pateman's words, "cannot be washed clean of the history of justification of subjection" (Pateman and Mills, *Contract and Domination*, 15). Secondly, this essay shall discuss three reasons why western philosophy is not irredeemably racist: the inquisitive nature of philosophy; the fact western philosophy has been used as a framework to develop liberating ideas for marginalised groups alongside those used to subjugate them; and the fact that some areas of western philosophy are indeed separate from the issue of race and racism. Thirdly, I turn my attention to how western philosophy may overcome its racism through a global approach that integrates historical context when examining philosophers and their ideas, to prevent the encroachment of discriminatory biases. Finally, this essay responds to an important counter-argument to the essay: how is this essay philosophy?

## II. PRELIMINARIES

Before continuing to the sections outlined above, it is important to highlight and clarify a few things. The titular question is a gargantuan one given both its scope and subject matter. Therefore, this essay will inevitably leave out areas and have weaknesses. However, just because it is a difficult question to answer does not mean it should not be a question that is asked, because it is still an important question. Thus, whilst this essay may be susceptible to certain criticisms, it is still vital to kickstart the process of answering the titular question.

It is also essential to clarify what is meant by both 'western philosophy' and 'irredeemably racist' given both are nebulous and vague terms. Firstly, by 'western philosophy' this essay means the types of philosophy commonly taught in academic institutions across the western world, e.g., in the UK, USA, and Continental Europe. This also includes what Dotson refers to as 'professional philosophy' (Dotson, "How is this Paper Philosophy?", 3). The scope of this essay should also be highlighted here. To fully answer the titular question would require a few hefty tomes to completely address whether every aspect of western philosophy is irredeemably racist. Instead, I will present a series of paradigmatic examples of western philosophy that brings out its racism. I am aware a critic may raise the objection of 'cherry-picking' philosophers who are not representative of western philosophy. In my view, they are indeed representative of much of western philosophy. Although to adequately respond to the cherry-picking criticism, it would require an in-depth discussion of many examples to show how the ones used in this essay are representative. Such a discussion is not the focus of this essay, thus properly responding to this claim is also beyond this essay's scope. However, what can be said is the fact that the specific examples of racism from philosophers do not necessarily matter. What instead matters is the fact that it is the ahistorical nature of western philosophy that means it ignores the racism of its philosophers, regardless of whom those philosophers may be. This essay will also address areas of philosophy that are separate from the issue of race and racism, such as logic. Thus, this essay understands not all of western philosophy can be subjected to the titular claim. Although as will be discussed below, even those parts of western philosophy that are separate, should adopt a new approach and examine historical context.

Secondly, when considering this notion of 'irredeemably racist' applied to

philosophy, we need to think about a few things. We must first establish what it means for western philosophy to be racist. Western philosophy being racist differs from, for example, an individual being racist. This is because unlike an individual's racism, exhibited on some physical level (e.g., verbal, or physical abuse), western philosophy's racism is exhibited on a metaphilosophical level. Western philosophy's racism is most clearly characterised by the fact its ideas (especially from political philosophy) have been used to subjugate those who are not white. However, an arguably more important and fundamental feature of western philosophy's racism is its evident inability to tackle racism within its works (despite its capacity to do so). Now that we have established by what it means for western philosophy to be racist, we can understand what is meant by the term 'irredeemably racist'. Something irredeemably racist is something beyond the point of redemption and ought to be abandoned altogether. Quite a clear example of something that is irredeemably racist is the KKK. This is because of the terror it has imposed on African-Americans over the last one hundred plus years. I propose three criteria for something to be *irredeemably* racist, i.e., beyond redemption. Most importantly, if something cannot firstly acknowledge and then challenge its racism, it is irredeemably racist, as without the capacity to challenge one's racism there can be no internal change away from racism and towards redemption. Thus, this would suggest that a criterion for western philosophy not being irredeemably racist is: if it has the capacity to question whether it is irredeemably racist. Something that is also beyond redemption in its racism *could* also not have been used to develop ideas that have liberated non-whites. To do otherwise would suggest some redemptive elements (namely those liberating ideas) to a particular object. Therefore, this would suggest that a second criterion for western philosophy not being irredeemably racist is: if it has been used as a framework to develop liberating ideas for ethnic minorities. Thirdly, an object that is irredeemably racist also suggests that all parts of that object are racist. Otherwise, if some parts of an object were not racist, then not all of the object can be deemed racist, let alone irredeemably racist. Subsequently, this leads us to derive our third condition for western philosophy not being irredeemably racist: if some areas of western philosophy are separate from the issue of race and racism. As an aside, it is at this point worth pondering the notion of systemic racism. The term is often used when describing an institution. For example, if someone were to call the BBC 'systemically racist' they would refer to how the BBC recruits employees, picks news stories, and so on. In other words, they would be referring to the system of operation at the BBC and how it is this system that is racist. Therefore, calling western philosophy 'systemically racist' would also require us to establish the system of operation of western philosophy. This in of itself is such a mammoth task and would require an essay to answer fully. Thus, even though calling western philosophy systemically racist may not be a bad claim given (as will be argued below), western philosophy is fundamentally ahistorical (i.e., ahistoricity is *part* of western philosophy's system of operation) and this ahistoricity leads to an inability to confront racism within western philosophy. The claim western philosophy is 'systemically racist' merely is beyond the scope of this essay. Therefore, the essay shall restrict its claim to the fact that western philosophy is racist.

### III. IS WESTERN PHILOSOPHY RACIST?

In this first section, I will argue that western philosophy is racist. This is mainly due to western philosophy's ahistorical nature and the racist nature of western political philosophy (a part of western philosophy). I will argue that the ahistoricity of western philosophy makes it racist by considering examples of paradigmatic western thinkers and showing how the lack of criticism of their racism results from the ahistoricity of western philosophy. I shall also argue that the racist undertones of modern political philosophy, particularly social contract theory, means at least part of western philosophy has been used to justify the subjugation of non-whites. Thus, making it racist.

The main reason why western philosophy is racist is because of its ahistorical nature. One way this ahistorical aspect of western philosophy demonstrates this racism is, via western philosophy's inability to acknowledge the racism of its most prominent thinkers and the influence their racist ideas has had. The three chosen Mill, Hume, and Kant have been considered because they have been exceptionally important to western philosophy and represent a large portion of it.<sup>96</sup> The most prominent thinkers have been explicitly chosen because if western philosophy ignores those thinkers' racism, it is also likely to ignore the racism of less notable thinkers. It should be noted here that there will first be an overview of their racism and its impact. However, the main reason their racism is being discussed is to illustrate a point. This point is that western philosophy's failure to confront the racism of its philosophers is a direct result of its ahistorical nature. Who the most prominent thinkers are specifically is not necessarily vital to the thesis of this essay. The main point of this section is to show it is the ahistorical nature of western philosophy that means it ignores the racism of its philosophers, regardless of whom those philosophers may be. Although when discussing Kant's racism, it is important to highlight the importance of his ideas in the invention of race itself.

J.S.Mill's racism<sup>97</sup> is most clearly demonstrated through his racism in *On Liberty*, where he states the following "Despotism is a legitimate mode of government in dealing with barbarians" (Mill, *On Liberty*, 23). Mill is essentially saying here that all the liberties he has spent the entire book arguing for should be afforded to those who are 'civilised' i.e. European whites, and not to those who are 'barbarians' i.e. non-white Europeans (Beate, "Barbarian Thoughts: Imperialism in the Philosophy of John Stuart Mill.", 599-618). For example, in *On Liberty* Mill argues for the Harm Principle. This is the idea that one should have their liberty regulated if they were going to cause harm to others. This principle was not afforded to 'barbarians', i.e., 'the uncivilised' non-whites whom the British colonised. Instead, their liberty was restricted because they did not know how to rule themselves and therefore 'despotism' was a legitimate mode of government. In this way, Mill's racism can be understood as an 'educational racism', i.e., whites need to teach non-whites how to govern themselves. Mill's educational racism is perhaps best encapsulated by B.Williams' reference to "Government House utilitarianism" (Williams and Moore,

---

<sup>96</sup> For example, Mill's Utilitarianism and Kant's categorical imperative have both been important to western moral philosophy. Hume's problem of induction and empiricism have also been deeply influential in western philosophy. These are just to name a few examples of how pervasive the ideas of Mill, Hume, and Kant have been in western philosophy.

<sup>97</sup> For further examples of J.S.Mill's racism please refer to Beate article referenced below.

Ethics and the Limits of Philosophy, 108). This phrase reflects the colonial elitism found within Mill's educational racism. All of this is to say that Mill's educational racism is an issue within his works, nonetheless an issue in one of his most famous works. As will be discussed later on, it is the ahistorical aspect of western philosophy which means it fails to confront Mill's educational racism, and it is the failure to confront this racism which makes it racist.

Hume's racism is best categorised as polygenesis racism.<sup>98</sup> This is racism that argues that people from different races are from different species of man. Those that are not white belong to an inferior species than those that are white. For example, in a footnote to "Of National Characters" Hume said the following:

"I am apt to suspect the Negroes to be naturally inferior to the Whites. There scarcely ever was a civilized nation of that complexion, nor even any individual, eminent either in action or speculation. No ingenious manufactures amongst them, no arts, no sciences. On the other hand, the most rude and barbarous of the Whites, such as the ancient Germans, the present Tartars, have still something eminent about them, in their valour, form of government, or some other particular. Such a uniform and constant difference could not happen, in so many countries and ages, if nature had not made an original distinction between these breeds of men." (Hume, "Of National Characters", 213)<sup>99</sup>

Hume's reference to "naturally inferior" and "breeds of men" shows that his racism lies clearly within the polygenesis camp. It should be noted that when considering the example of Hume, his racism differs from that of Mill and Kant. As unlike those two, his racism is not important to his philosophy. One of Hume's fundamental philosophical points was that he was an empiricist. Indeed, this empiricism, and the probabilistic inductive claims it leads to, is arguably reflected in his racism. For example, in the above passage, he says, "I am apt to *suspect* the Negroes to be naturally inferior to the Whites." Subsequently, with this example of Hume it is important to consider not so much the impact of a philosopher's racism on their philosophy, but rather the impact of their racism on the world. Indeed, whilst Hume himself denounced slavery on the basis of its immorality, it is undeniable that his kind of polygenesis racism (which Hume along with other enlightenment philosophers were associated with) that he was advocating for, was used by those in favour of slavery in Antebellum America to justify the enslavement of black people (Luse, "Slavery's Champions Stood at Odds: Polygenesis and the Defense of Slavery.", 383). For example, in 1844, US Secretary of State Calhoun wrote a letter to then Foreign Secretary of Britain, Lord Pakenham. Calhoun protested British attempts to get Texas to emancipate its slaves arguing from a polygenesis perspective. For example, Calhoun wrote that cranial measurements of blacks, an important technique for polygenists, showed their inferiority and thus suitability for slavery (Luse, "Slavery's Champions Stood at Odds: Polygenesis and the Defense of Slavery.", 385). This, therefore, shows that polygenesis racism, which Hume advanced, was then used in the 19<sup>th</sup> century to justify slavery. What

---

<sup>98</sup> For a more detailed exploration of Hume's racism please refer to Ten, C.L. "Hume's Racism and Miracles." *The Journal of Value Inquiry* 36, no. 1 (2002): 103-09.

<sup>99</sup> Also see T.H. Huxley, *Hume* (London: Macmillan, 1902), pp. 22–24.



is even more damning is that, as Immerwahr observes, Hume's above quotation was "widely quoted by racists and defenders of slavery." (Immerwahr, "Hume's Revised Racism.", 481). This, therefore, shows that Hume's racism had a clear and measurable impact on the world, given it was used by racists and those in favour of slavery.

Before turning to how the ahistorical nature of western philosophy means it ignores its philosophers' racism, which in turn makes it racist, the most vile and arguably influential racism of our last thinker needs to be considered, namely Kant's racism. Kant's pseudo-scientific anthropological racism can be classified as 'transcendental racism'. This is because, through his anthropology and theory of race, he argued that what struck at the heart of an individual's personhood was their whiteness. Thus, if they lacked such whiteness, they were considered sub-human. Whilst the precise details of Kant's anthropological racism will not be discussed here fully; he did, for example, argue for a racial hierarchy: white, black, yellow and red (Zhavoronkov and Salikov, "The concept of race in Kant's Lectures on Anthropology.", 275-290). He also believed that all individuals were born white and that their environment tainted their skin (Zhavoronkov and Salikov, "The concept of race in Kant's Lectures on Anthropology.", 280). Regardless of the details of Kant's racism, it may be asked how his moral philosophy, the categorical imperative (which argued "Act in such a way that you treat humanity, whether in your own person or in the person of any other, never merely as a means to an end, but always at the same time as an end."), can be reconciled with a statement such as "All races will be extinguished ... only not that of the Whites." (Kant and Ellington, *Groundwork Of The Metaphysics Of Morals*, 36) (Mills, *Black Rights/White Wrongs*, 97). My interpretation is that Kant's moral philosophy does not see non-white people as human. Kant did not need to explicitly say this in his moral philosophy because his transcendental racism informed his perspective on what counts as being human, i.e., whiteness. In a similar fashion to how the framers of the US constitution talked about the "Life, Liberty and the pursuit of Happiness" which were "instituted among Men", i.e., the white man, by humanity Kant meant white humans (Jefferson, *Copy of the Declaration of Independence*). Although, this is my own interpretation of Kant there are many ways to interpret Kant so that his racism did not inform his moral philosophy. One such way is by saying that Kant was not consistent across his ideas. Therefore, his racism, which would exclude non-whites from the categorical imperative, may not have informed his moral philosophy. My own response to this is that Kant himself probably thought his ideas were consistent with one another. This debate is, however, almost irrelevant to the thesis of this essay. Perhaps, Kant's racism did inform his philosophy, perhaps it did not; it is not a crucial point. What matters to the thesis of this essay, is that (as will be argued below) by abandoning western philosophy's ahistorical nature and allowing for the examination of historical context, we can understand whether Kant's racism did or did not inform his philosophy. We need a holistic view of Kant to make a proper judgement.

In addition to how Kant's racism may have informed his philosophy, we need to consider the impact of Kant's racism. Out of the three thinkers discussed Kant's racism is certainly the most influential, what is more, Kant's racism can arguably be considered fundamental to our modern conception of race. As R. Bernasconi argued in his now-famous essay "Who invented the concept of race?" he argues that it was Kant (Sandford, "Kant, Race, and Natural History", 951). To be clear by

‘inventor of race’ it is meant “the one who gave the concept sufficient definition for subsequent users to believe that they were addressing something whose scientific status could at least be debated” (Sandford, “Kant, Race, and Natural History”, 951). This claim holds significant weight when considering that Kant’s theory of race was a profoundly influential contribution to the late 18th-century German life sciences. For example, it was adopted by scientists with much more scientific credibility than Kant, like Christoph Girtanner and Hans Blumenbach (Sandford, “Kant, Race, and Natural History”, 951). Therefore, this means Kant’s theory of race helped advance the essentialist, biological theories of race that have been instrumental to European racism since the late 18th century. What is more, some scholars such as C.Mills argue that Kant’s distinction between personhood and sub-personhood was later drawn upon by Nazi racial theory (Mills, *The Racial Contract*, Chapter 2). Thus, although there may be a scholarly debate about the extent to which Kant can be classified as being the inventor of our modern conception of race, what is clear and undeniable is that Kant’s racism had a clear impact on the world.

Now that the racism of our prominent thinkers and the impact of their racism have both been explained. It is important to consider how it is the ahistorical nature of western philosophy that prevents it from confronting its thinkers’ racism. By ahistorical, it is meant that western philosophy lacks any historical perspective. Western philosophy is ahistorical because it rejects examining the historical context, arguing that to do otherwise is beyond philosophy, instead it prefers to stay within a bubble of ideas.<sup>100</sup> The precise reason western philosophy rejects historical context will be discussed when considering a critical counter-thesis to this essay: how is this essay philosophy? For now, all that needs to be understood is how it is this rejection of historical context that means western philosophy ignores the racism of its thinkers. If for example western philosophy did incorporate historical context about Kant (i.e. context about his racism and how it informed his philosophy) into literature and teaching it would be finally able to confront the racism within his works. Therefore, since western philosophy’s ahistorical nature prevents it from looking at its philosophers’ racist context, of its philosophers it is this which makes it racist.

This essay will now turn to two typical responses from western philosophy that arise when bringing up philosophers’ racism. The following counter-arguments that will be discussed can be summarised by a ‘so what’ attitude towards philosophers’ racism in general. They both accept that there is indeed racism in western philosophers’ works, but we should ignore that racism.

Firstly, there is a response that typically goes along the lines of “aren’t these thinkers and their racist views just products of their time?” Whilst there may be some truth in the statement, in so far, that our environment can inform and influence our views. The impact of our environment on our views, does not mean that we should not examine their racism and the influence it had on the world. Secondly, this view presupposes that their racism will not be of importance to understanding their philosophy, which may not hold true in the case of Kant. Thirdly, to say that these thinkers were products of their time discredits those

---

<sup>100</sup> By this it is meant that western philosophy is only concerned with the ideas of philosophers. Western philosophy is not concerned with why philosophers may have developed certain ideas. It prefers to stay referring to platonic ideas rather than looking at the historical context of those ideas.

who did not think like them. For example, take Hume and his polygenesis racism. His racism was opposed by James Ramsey, an abolitionist, who said Hume's racism was "made without any competent knowledge of the subject." (Ten, "Hume's Racism and Miracles, 103-109). Therefore, on the above three accounts, we can reject this first typical 'so what' response from western philosophy.

The second common 'so what' response typically put forward by western philosophy goes along the lines of "isn't their racism separate from their philosophy?" As already shown with the likes of both Mill and Kant, their racism could inform some of their philosophy. It may be responded, "can't we just separate their racism from their philosophy?". For example, if for argument's sake, Kant's racism did inform the categorical imperative, can we not just reinterpret Kant's categorical imperative to mean all humans (regardless of race or ethnicity) when we use the word 'humanity'. Doing so would seemingly show that we can separate racism from the ideas of racist philosophers. Indeed, we can do this; we need not completely abandon an idea of great philosophical significance that is also tainted by racism. However, when doing this, it is important to acknowledge the historical context of the idea, the racism of the philosopher who developed it, and how their racism may have influenced their philosophy. Moreover, such a separation may require a renaming of ideas. If, for example, we separate the categorical imperative from Kant and his racism, it would no longer be apt to call it Kant's moral philosophy as by 'Kant's moral philosophy' we now mean the ethical system which is informed by Kant's transcendental racism. All of this is to say that the two 'so what' arguments fail to show that philosophers' racism is irrelevant to philosophical inquiry.

A second reason why western philosophy can be considered racist is because of the impact of western political philosophy and how it has been used to justify subjugation. It should be noted that unlike the above reason about the ahistorical nature of philosophy, the discussion of this reason shall be limited in this essay. This is mainly because it has been thoroughly explored by the likes of C.Mills and C.Pateman in their works. However, it is still important to acknowledge how behind social contract theory, which has been the basis for much of modern western political philosophy, lies the 'Racial Contract'. The racial contract is, as Mills argues, the set of agreements/meta-agreements between one subset of humanity (i.e., white people), which is then imposed on everyone else excluded from that determined subset (Mills, *The Racial Contract*, Chapter 1). This is done to exploit the differentiated group (i.e., non-whites) to extract resources, land, and deny them socio-economic opportunities. Moreover, as Mills argues, the racial contract has underwritten social contract theory. For example, Hobbes saw native Americans as the real-life example of the savages he envisioned that would exist in his state of nature (Mills, *The Racial Contract*, Chapter 2). Similarly, Locke saw non-whites as having inferior intellect in leaving the state of nature at a slower pace than whites. Therefore, to him, whites were free to exploit the resources and land of non-whites (Mills, *The Racial Contract*, Chapter 2). Moreover, for Rousseau, the only natural savages are those who are non-white. Even if Rousseau did indeed call native Americans the "noble savage", as Mills rightly points out, they are still savages, i.e., primitive sub-humans excluded from society (Mills, *The Racial Contract*, Chapter 2). These three examples, therefore, help demonstrate that the racial contract has underpinned social contract theory. Mills does, however, argue that social contract theory can be reworked within a Rawlsian

framework. Therefore, to Mills, social contract theory can rectify historical racial injustices and racial discrimination, instead of perpetuating them. However, to some such as Pateman, social contract theory “cannot be washed clean of the history of justification of subjection” and therefore should be abandoned (Pateman and Mills, *Contract and Domination*, 15). Whilst there may be a debate about whether social contract theory is irredeemably racist (or in Pateman’s case irredeemably sexist). What is clear is that the central concept in modern western political philosophy, the social contract, has been used to justify the subjugation of non-white people. Therefore, we are justified in saying that western political philosophy, and thus part of western philosophy, is racist.

Before concluding this section, it is worthwhile considering the impact of ancient western philosophers and their political philosophy. For example, in *The Politics*, Aristotle’s teleological philosophy denoted some as “slaves by nature” (Aristotle, Sinclair, and Saunders, *The Politics*, 69). This is because Aristotle saw the relationship between master and slave conforming to the natural binary pattern he saw. This included distinctions such as better/worse, male/female, rational/irrational, and ruler/ruled. He also thought that the justification of slavery lies with the “moral superiority” of the master. As to Aristotle, coerced slavery is just if and only if those coercing are morally superior. It is worth highlighting that the slaves and enslavement Aristotle spoke of normally applied to those captured in war. It may then be objected, that although Aristotle’s political philosophy is reprehensible, it does not have to do with the subjugation of non-white races given such a distinction of races did not exist in ancient Greece. This is indeed true, however as T.A.Sinclair notes “arguments which [Aristotle] used were still used among the defenders of slavery in the nineteenth century” (Aristotle, Sinclair, and Saunders, *The Politics*, 21). This is because in differentiating between black and white, enslavers also imposed Aristotle’s binary view of nature. What it is more, even if Aristotle’s political philosophy had not been used to justify racism, his ideas still fed into the justification of subjugation. All of this is to say that in addition to modern western political philosophy, ancient western political philosophy has also been used to justify the subjugation of non-white people, just not by the ancient philosophers who developed those ideas.

This section has argued that western philosophy is racist. Firstly, western philosophy’s ahistorical nature means it ignores the racism of its philosophers and the impact of their racism. Secondly, western philosophy can be and has been used for racial subjugation. For example, western political philosophy and the social contract. Although admittedly a less important reason given, it only applies to western political philosophy. It is for these two reasons that philosophy can be considered racist.

#### IV. IS WESTERN PHILOSOPHY *IRREDEEMABLY* RACIST?

Although western philosophy’s ahistorical nature and the impact of western political philosophy make western philosophy racist, it is important to remember that does not translate it into being irredeemably racist. This is because, for something to not be irredeemably racist, not beyond redemption, it must (as argued at in the introduction of this essay) fulfil the following three sufficient conditions:

- 1) If it has the capacity to question whether it is irredeemably racist.

2) If it has been used as a framework to develop liberating ideas for ethnic minorities.

3) If some areas of western philosophy are separate from the issue of race and racism.

As will be discussed in the following section, western philosophy is successful on all three accounts; therefore, it is wrong to suggest that western philosophy is irredeemably racist, even if it is racist.

The first and most important reason western philosophy is not irredeemably racist is because of its inquisitive nature, which means western philosophy has the capacity to question whether it is irredeemably racist. A central tenet of philosophy is always questioning the nature of our reality, always asking the question of why and questioning assumptions. This inquisitive norm equips philosophy with the tools to give it the capacity to overcome its racism. Redemption begins with self-reflection. To redeem oneself from some past error, the first thing is to realise, acknowledge and accept that one made the error in the first place. This can only be done when one reflects upon itself. Luckily, philosophy is just the kind of discipline which excels at self-reflection, given its inquisitive nature. After all, how would philosophy question its foundations, and the racism within it, without this inquisitive engine? The inquisitive nature of philosophy is also arguably a common justifying norm of philosophy, which, as will be discussed below, is ultimately the reason why even philosophy's capacity also includes this article as philosophy. Moreover, this can be regarded as the most important reason because, unlike the next two reasons, it allows us to question both the ahistorical nature of philosophy and the impact of western political philosophy. Put simply; it is the inquisitive nature of philosophy that gives us the capacity to question the racism of western philosophy, whereas the other two reasons respectively demonstrate some redemptive qualities of western philosophy and areas of western philosophy separate from the issue of race and racism. This is all to say that western philosophy is successful in regards to the first condition outlined above.

In addition to the inquisitive nature of philosophy, the simple fact that western political philosophy has been used as a framework to develop liberating ideas for non-white people means it is successful in regards to the second condition. This, in turn, makes the position that western philosophy is irredeemably racist even more untenable. For example, one only needs to look to those who opposed slavery based on the 'natural rights' that Locke references (Locke, *Second Treatise of Government*). Or indeed those like C.Mills who reconstructs a Rawlsian social contract that also factors racial injustices (Mills, *Black Rights/White Wrongs*, Chapter 9). This, therefore, demonstrates western political philosophy, a subset of philosophy, has indeed been used as a framework to develop liberating ideas for ethnic minorities. Since western political philosophy is a subset of philosophy, the same can also be said of western philosophy. Subsequently, since western philosophy includes ideas that have been used for both the subjugation and liberation of non-white people, it is mistaken to call it irredeemably racist, beyond redemption, because there are some redemptive qualities (namely the liberating ideas developed). One need not expound on this reason too much given only a few examples are necessary to successfully fulfil the second condition of western philosophy not being 'irredeemably' racist.

The final reason why western philosophy cannot be considered irredeemably racist is that some areas are indeed separate from the issue of race and racism. The clearest example of an area of western philosophy completely separate from race and racism is logic. This because unlike say Kant's moral philosophy, the mathematical nature of logic means it has no concern for the social world. Thus, it is completely separate from the issue of race and racism. This, therefore, fulfils our third condition, allowing us to resoundingly reject claims that western philosophy is irredeemably racist. However, it may be argued that if we do indeed say that some areas of philosophy are separate from racism. Then are we not also saying that the racism of philosophers whose ideas are separate from race can be considered irrelevant. For example, take the anti-semitism of Frege, a famous logician (Monk, "Gottlob Frege: The Machine in the Ghost"). Since his anti-semitism<sup>101</sup> had nothing to do with his logic, are we not in essence, saying that his racism should be ignored? This would then seemingly take us back to, in logic at least, this notion of ahistoricity in western philosophy which this essay is arguing we should abandon. No, the racism of Frege and all philosophers still matters, given the impact it has on our social environment. Just not necessarily to their philosophy though. Regardless, in making a judgment of how much a philosopher's racism informs their philosophy or vice-versa, having a greater understanding of historical context allows us to judge whether or not their biases taint their philosophy. Therefore, even in those areas of western philosophy separate from the notion of race and racism, we should still abandon ahistoricity and examine historical context.

This section has argued that although it is correct to designate western philosophy racist due mainly to its ahistoricity and the impact of western political philosophy. It is incorrect to refer to western philosophy as irredeemably racist. Firstly, the fundamentally inquisitive nature of philosophy allows us to question the ahistorical nature of it and how this informs its racism. Therefore, since western philosophy has the capacity to question its racism, it is not irredeemably racist. Secondly, although western political philosophy, and thus western philosophy, has created ideas that were used for the justification of subjugation, it has also had very liberating ideas developed within its framework. Therefore, since western philosophy has some redemptive qualities, it cannot be regarded as irredeemably racist. Thirdly, as discussed above some areas of western philosophy are separate from the issue of race and racism. Therefore, they cannot be judged as racist, let alone irredeemably racist. Thus, not all of western philosophy can be deemed as irredeemably racist. However, it is still important to learn the historical context of thinkers and abandon ahistoricity. As by doing so, we as philosophers are better equipped to understand how an individual's biases may have informed their philosophy, or vice-versa.

## V. PHILOSOPHY IS STILL RACIST. WHAT SHOULD WE DO ABOUT IT?

Although the previous section established that western philosophy is not irredeemably racist, as established in the first part, western philosophy is still racist. Therefore, the question remains, what are we to do about the racism within western philosophy? Here I will consider two previous attempts at answering this question and argue that they are inadequate. I will then propose and defend my own possible solution.

Firstly, there is Van Norden's multicultural approach. He argues that western philosophy's apathetic and often exclusionary attitude to philosophical ideas outside

<sup>101</sup> For example, in Frege's Nachlass, his diary, he was deeply opposed to Catholics, the French and, above all, Jews, who he thought ought to be deprived of political rights and expelled from Germany.

the western canon is down to “racial nationalism” (Van Norden, *Taking Back Philosophy : A Multicultural Manifesto*, 88). To Van Norden “the desire to draw a sharp boundary between Anglo-European philosophy and supposedly non-philosophical thought is a manifestation of a broader pattern of xenophobic, chauvinistic, nationalistic, and racist efforts to separate ‘us’ from ‘them’.” (Van Norden, *Taking Back Philosophy : A Multicultural Manifesto*, 84) Van Norden’s solution to overcoming this “racial nationalism” is to adopt a multicultural approach in western philosophy (Ganeri, “*Taking Philosophy Forward*”). For example, this would translate into western philosophy departments that offer teaching and research into many philosophical cultures. These cultures would then be “brought into dialogue” to place no one philosophical culture above another. Such an approach would allow western philosophy to overcome its “racial nationalism” and broaden western philosophy’s diversity.

However, Van Norden’s multicultural approach also raises fundamental issues for western philosophy. Whilst, Van Norden is undoubtedly right in suggesting that western philosophy needs to look to other philosophical traditions to overcome racism. As Ganeri argues we do not solve the issue of ethnocentrism (not racism) by multiplying cultures in dialogue but rather instead subtracting away this manufactured notion of ‘culture’ and being open to philosophical ideas from anywhere (Ganeri, “*Taking Philosophy Forward*”). This is because of the ‘embrace and fragment’ or ‘exclude and contract’ dilemma faced by western philosophy if it took Van Norden’s multicultural approach. For example, under Van Norden’s dichotomous approach, we ought to either embrace multiculturalism or not. If we did indeed embrace other philosophical cultures in their entirety, western philosophy might become more concerned in how the traditions have a dialogue, rather than answering philosophical questions. Would it not eventually make the phrase ‘western philosophy’ meaningless? Would it not be more apt to call this multicultural philosophy ‘comparative philosophy’ rather than western philosophy? Although this essay is indeed arguing for a more global approach to western philosophy, it is not saying that we should adopt a more comparative approach. Moreover, highlighting Van Norden’s multicultural approach’s unfeasibility is also the practical issue of a severe lack of academics around the globe who would be able to fill western philosophy departments to ensure a truly multicultural approach is taken everywhere. This is because as Ganeri observes it is not necessarily the “racial nationalism” of western philosophers that makes them adverse to the works from other philosophical traditions, but rather their ignorance. On the flip side, if we do not embrace Van Norden’s multicultural approach, we maintain the exclusionary course that western philosophy is currently on. Therefore, the way to circumvent this issue is to, as Ganeri suggests, take a more global rather than multicultural approach, which is open to the genius of philosophical ideas from anywhere rather than trying to incorporate every philosophical culture. The multicultural approach is concerned with classifications such as Islamic, Chinese, or African philosophy.<sup>102</sup> As Ganeri argues, by removing this notion of ‘culture’, western philosophy will not

---

<sup>102</sup> Simply put Van Norden’s multicultural approach is one that places an emphasis on differing philosophical traditions such as the Islamic, African, and Chinese philosophical traditions. Van Norden then advocates that these differing traditions be brought into dialogue with one another on various issues in philosophy, such as the mind-body problem. Such an approach is, for example, less concerned with answering the mind-body problem and is instead concerned with how the different traditions interact with one another. This is in stark contrast to Ganeri’s global approach which removes this notion of tradition and culture. The global approach instead looks at philosophical ideas regardless of which philosophical tradition they emerged from and integrates them into the western framework. The global approach, unlike the multicultural one, is concerned with, for example, answering the mind-body problem using non-western ideas. It is not concerned with how the various philosophical traditions, from which ideas emerged, interact with one another.

‘embrace and fragment’ because it will be concerned with non-western ideas and how they interact with other ideas in western philosophy. Thus, western philosophy would not become the comparative philosophy that would result from the dialogue between a multitude of philosophical traditions.

It should be noted that Ganeri’s approach is not without issues. Firstly, Ganeri argues that it is not racism but rather ethnocentrism that is the issue that needs solving. However, western philosophy is racist (for reasons discussed above). Moreover, Ganeri argues that western philosophy is not racist because he sees no difference in commitment to racist ideologies in western philosophy than in other branches of the humanities. However, he identifies the wrong issue. It is not the ‘commitment’ racism but rather western philosophy’s clear inability to confront racism within its canon and the influence of said racism. This inability of western philosophy is mainly due to the ahistorical nature of western philosophy. Moreover, there remains a significant issue in both the approaches of Ganeri and Van Norden. This is because integrating either other philosophical cultures or philosophical ideas from around the globe leaves us open to potentially integrating more biases (from those philosophical traditions found around the globe) within western philosophy. Therefore, both solutions would seemingly take us back to square one and may make philosophy more racist/discriminatory rather than less so. Therefore, this is why, as has been argued throughout this essay, examining historical context of philosophers alongside their ideas is key. For example, imagine we are examining an idea from Islamic philosophy, much like Kant’s categorical imperative it may be tainted by some discriminatory bias derived from the philosopher who thought of the idea, which may be sexism. By looking at the historical context of that philosophical idea, we would be better equipped to understand how the sexism of that philosopher informed his philosophical ideas or vice-versa.

This section has shown that the solution to western philosophy’s racism lies in a global approach that, as Ganeri argues, incorporates philosophical genius regardless of the philosophical culture from which it emerged. Van Norden’s multicultural approach leads to a dilemma between ‘embrace and fragment’ or ‘exclude and contract’. However, Ganeri is mistaken in arguing there is no racism within western philosophy because it has no more ‘commitment’ to racist ideologies. It is not the commitment to racism, but rather the inability of western philosophy to confront said racism that makes it racist. Furthermore, if we take Ganeri’s global approach it is even more important that we abandon the ahistoricity of western philosophy and examine historical context to prevent the incorporation of further biases into western philosophy.

## VI. HOW IS THIS ESSAY PHILOSOPHY?

Finally, I shall turn to an important criticism of this essay: how is this essay even philosophy? An objector might say: this essay, with its call for historical engagement, does not constitute philosophy. But rather, such means of reflection are actually parts of history and sociology. I wish to conclude by defending against this objection by arguing that this essay does subsume under the concept of the ‘questioning norm’ of philosophy. The question (how is this essay philosophy?) is one that places a value on an exclusive legitimization narrative that fails to fully encapsulate what is really the common justifying norm of philosophy. This



exclusive legitimation narrative is one that in essence, says ‘that in order to be classified as western philosophy, one’s ideas must be congruent with a traditional perception of what counts as philosophical engagement’. The exact details of what counts as this ‘traditional perception’ will not be discussed here given, as Dotson points out, it is entirely contingent on the question-asker (that is of ‘how is this essay philosophy’). However, what does matter is that this legitimation narrative is one that is exclusive. For example, G.Salmon argues that the very act of making congruent runs counter to the queer method. This is because Salmon’s queer method begins at a point of estrangement that is already at the opposite side of being congruent (Dotson, “How is this Paper Philosophy?”, 6). This, therefore, shows that this specific type of culture of justification found within western philosophy is itself exclusive to those ideas that fall outside what can be seen as a traditional perception of philosophy. According to this exclusive legitimation narrative, this essay may very well be considered to be beyond western philosophy because it examines western philosophers’ racist history, i.e., something beyond the traditional conception of what counts as philosophy. It is this type of culture of justification that underlies western philosophy’s ahistoricity. This is because examining historical context is seen as lying outside the realm of typical culture of justification associated with western philosophy. It is seen as something which belongs to the realm of history and sociology, rather than philosophy.

However, if we as Struhl argues, adopt a more basic and much more fundamental common justifying norm for what counts as philosophy, then examining historical context can indeed be part of philosophy. This is because to Struhl what can be considered philosophical engagement ultimately comes down to the “critical and systematic investigations” of fundamental assumptions (Dotson, “How is this Paper Philosophy?”, 10). The only way one can accurately assess our fundamental assumptions of how philosophers and their ideas are separate from the idea of race is by examining those philosophers’ historical context. Therefore, under this much more basic and universal norm, which will be called the ‘questioning norm’ henceforth, western philosophy ought to integrate historical context.

It is here that I differ with Dotson, given she wants to shift away from a culture of justification in philosophy towards a culture of praxis. When considering Struhl’s argument for the questioning justifying norm, she argues that his argument only holds value if this questioning norm is “taken to be univocally relevant to all philosophical enterprises.” (Dotson, “How is this Paper Philosophy?”, 10). To Dotson, Struhl’s argument rests on the fact that such norms for philosophy such as the questioning norm even exist and are relevant to all of philosophy. However, Struhl’s “critical and systematic investigations” of *fundamental assumptions* does to my mind seem to be the most basic philosophical norm that there is. Therefore, Struhl’s questioning norm can be seen as something that both exists and is relevant to all of philosophy. This inquisitive and questioning norm gives western philosophy both the capacity to include this essay as philosophy and the capacity to question its racism, thus making it not irredeemably racist. In answering the objection (how is this essay philosophy?) we take the first step towards the metaphilosophical project that the essay advocates for, recognising racism within western philosophy. As by understanding that historical engagement can be included as philosophy, we can begin to acknowledge and engage with the racism found within in the western canon.

The answer to whether western philosophy is irredeemably racist is no; it is not *irredeemably* racist. Western philosophy is, however, still racist. This is mainly due to its ahistoricity, which means it is unable to confront the racism of its philosophers and the impact their racism had. But also because of the way western political philosophy has been used to justify the subjugation of non-whites. However, whilst western philosophy is racist, it is not irredeemably racist. This is because there are three areas of redeemability to western philosophy. Firstly, but less importantly, western philosophy has areas that are separate from racism. Although admittedly even in such areas, the examination of historical context is still key to knowing whether a philosopher's biases inform their ideas. Secondly, and more importantly than the previous reason, is the fact that western philosophy, and the framework western political philosophy provides, has been used to develop liberating ideas. However, this is somewhat offset by the fact that western political philosophy has been used to justify the subjugation of non-white people, which is a reason why western philosophy can be considered racist. Finally, the questioning norm, highlighted by Struhl, is the most important reason why western philosophy is not irredeemably racist because it gives western philosophy the capacity to question its ahistorical nature. The solution to this racism within western philosophy is, as Ganeri argues, the incorporation of philosophical ideas regardless of the philosophical 'culture' they belong to. This will enable western philosophy to shift away from the racism found within it. Although, as noted before, it is necessary to incorporate the historical context when adopting this global approach. Firstly, to confront the racism of western philosophers. Secondly, to prevent the incursion of other potential biases that may emerge when we integrate non-western philosophers and thus take western philosophy back to square one. Whilst this global approach does indeed prevent the conceptual stretching of western philosophy so that it is still a meaningful term (unlike Van Norden's comparative multicultural approach which may 'fragment' the term western philosophy). It is important to understand that this does not mean the meaning of 'western philosophy' will remain static. This is because it will necessarily change to some degree as there is a greater integration of non-western philosophers. However, given the increasingly diverse backgrounds of people that choose to now study western philosophy, it will make the term 'western philosophy' a more representative and fitting one. Finally, I have argued that this essay can be defended against claims that it is not philosophy. Because the most basic and universal justifying norm, the questioning norm, means this essay can be included as bona fide philosophy.

BIBLIOGRAPHY

Allais, Lucy. "Kant's Racism." *Philosophical Papers (Grahamstown)* 45, no. 1-2 (2016): 1-36.

Aristotle, Sinclair, T. A, and Saunders, Trevor J. *The Politics*. Rev. Ed. / Revised and Re-presented by Trevor J. Saunders. ed. Penguin Classics. London: Penguin, 1992.

Cohen, Alix. *Kant's Lectures on Anthropology : A Critical Guide*. Cambridge Critical Guides. Cambridge, 2014.

Ganeri, Jonardon. 2018. "Taking Philosophy Forward - Los Angeles Review Of Books". *Los Angeles Review Of Books*.  
<https://lareviewofbooks.org/article/taking-philosophy-forward/#!>

Hume, David. "Of National Characters," *Essays Moral, Political and Literary* (London: Oxford University Press, 1963), p. 213, n. 1. Also see T.H. Huxley, *Hume* (London: Macmillan, 1902), pp. 22–24.

Immerwahr, John. "Hume's Revised Racism." *Journal of the History of Ideas* 53, no. 3 (1992): 481-86. Accessed December 26, 2020. doi:10.2307/2709889.

Jahn, Beate. "Barbarian Thoughts: Imperialism in the Philosophy of John Stuart Mill." *Review of International Studies* 31, no. 3 (2005): 599-618. Accessed December 21, 2020. <http://www.jstor.org/stable/40072091>.

Jefferson, Thomas. et al, July 4, Copy of Declaration of Independence. -07-04, 1776.

Manuscript/Mixed Material. <https://www.loc.gov/item/mtjbib000159/>.

Kant, Immanuel, and James W Ellington. 1993. *Groundwork Of The Metaphysic Of Morals*. 3rd ed. Hackett. Pg. 30

Locke, John. 1980. *Second Treatise of Government*. Edited by C. B. Macpherson. Cambridge, MA: Hackett Publishing.

Luse, Christopher A. "Slavery's Champions Stood at Odds: Polygenesis and the Defense of Slavery." *Civil War History* 53, no. 4 (2007): 379-412.

Mill, John Stuart. *On Liberty*. Cambridge Library Collection. Philosophy. Cambridge, 2012. Mills, Charles W. *The Racial Contract*. Ithaca, NY, 2014. Monk, Ray. 2017. "Gottlob Frege: The Machine In The Ghost". *Prospect Magazine*, , 2017. <https://www.prospectmagazine.co.uk/philosophy/the-machine-in-the-ghost>.

Pateman, Carole., and Charles W. Mills. *Contract and Domination*. Cambridge: Polity, 2007.

Sandford, Stella. "Kant, Race, and Natural History." *Philosophy & Social Criticism* 44, no. 9 (2018): 950-77.

Ten, C.L. "Hume's Racism and Miracles." *The Journal of Value Inquiry* 36, no. 1 (2002): 103-09. Van Norden, Bryan W. *Taking Back Philosophy : A Multicultural*

*Manifesto*. New York, 2017.

Williams, Bernard, and A. W. Moore. *Ethics and the Limits of Philosophy*. London ; New York: Routledge, 2006.

Zhavoronkov, Alexey and A. Salikov. "The concept of race in Kant's Lectures on Anthropology." *Kantian* (2018): 275-292.

---

---

# Stoic Love: A Senecan Defense of Friendship

---

---

*Gabriel Sánchez Ainsa*  
*University of Chicago*

## FOREWORD

*Editor Samuel Vucic*

What qualifies a *true*, or “Senecan” friendship, namely what makes a friendship philosophically justifiable?

In “Stoic Love: A Senecan Defense of Friendship”, Gabriel Sánchez Ainsa posits that the “Senecan” friendship instantiates the social nature of humanity, while simultaneously promoting in both friends mental tranquility, the good and happy life, and ethical projects for the sake of personal integrity and peace. In calling on Seneca’s first-century “CE Letters on Ethics to Lucilius,” Ainsa first delineates the relevant details of the seminal Stoic thinker’s ethical system as one aimed at cultivating a “bona mens” or “good mind” whose contents stand in proper relation to the external world, then proceeds to place *true* friendship within it. Rejecting Epicurean notions of self-interested and utilitarian friendship, Ainsa ultimately presents a comprehensive account of the authentic friend as not simply your benefactor, but your fellow steward of a shared moral vision for the world.

We selected this paper owing to the sheer and utter beauty of its prose and intrinsic significance of its subject matter. Argumentative validity and clarity surely abound within its pages, but this can be said for any submission which surpasses even the first round of our highly rigorous editorial process. This paper has something more, something essential, eternal, and even poetic. Ainsa’s work ever so eloquently brings the reader into an almost flow-like state of being, as they find themselves detached from time, lost in memories of their own friendships, pondering social life, love, and the pursuit of eudaimonia.

In recent times of great social distance and isolation, the importance of genuine connection and togetherness in friendship have become perhaps more pronounced than ever before. I now invite you, dear reader, to take a deep dive into the ancient wisdom which has helped me so profoundly to understand this importance, and be inspired by it within my own life.

## ABSTRACT

Friendship and love have a history of being treated as serious problems in the philosophical traditions, but often go unmentioned in contemporary literature. Still, the place that our friends and loved ones take in our everyday lives is awe-inspiring. How can philosophy explain the value of friendship? Is love worth it? To provide answers to these questions is the aim of this paper. To do so, I turn to a philosophical text committed to eudaimonistic ethical theory, a philosophy of a good and happy life, on the one hand and to the practices of friendship on the other, Seneca’s first-century CE *Letters on Ethics to Lucilius*. In this essay, I shall engage with Seneca’s text and philosophy to see what account of friendship he can offer and what reasons he delivers for the choice to share our lives with another person. I do not defend all the views presented by Seneca but treat him as an interlocutor; nor do I merely give an interpretation of Seneca’s views on friendship, but make a case for a sort of friendship that can be philosophically justified, which I happen to call “Senecan friendship.” In summary, I argue against the theses that human beings are not social by nature, that friendship is for the

sake of well-being, and that it is a threat to our ethical projects. Instead, I defend a Senecan friendship that is for the sake of a good life of mental tranquility and human virtue and does not threaten our other projects of personal integrity and peace; on the contrary, a good life asks of us that we expand our ethical projects beyond our individual selves.

## I. INTRODUCTION

Achilles stands up to rage for the death of his Patroclus; somebody somewhere decides to propose to her partner; I choose to move in with this and that friend. I believe that deep philosophical issues hide beneath these familiar scenes. For anyone who has ever loved has proclaimed that their loved one contributes to their happiness, has genuinely thought them good, and has had to figure out how to get to truly know their friend, to treat them well, to rightly act out of passion for them, to face their loss. Is all of this worth it? Why should we care for it? It is, at the end of the day, a problem of value. In a nutshell, to ask why we love our friends is to ask what is good for a human being.

To grasp these meaningful scenes of Achilles, the anonymous lover, and myself, we turn to a philosophical text committed to eudaimonist ethical theory, a philosophy of a good and happy life, on the one hand and to the practices of friendship on the other, Seneca's *Letters on Ethics to Lucilius*. This innovative Roman Stoicism of Seneca, manifested through epistolary conversation with his friend and student Lucilius, promises to offer a unique theory of love and friendship. In this essay, I shall engage with Seneca's text and philosophy to see what account of friendship he can offer, what reasons he delivers for the choice to share our lives with another person, and whether this Stoic love is one we can recognize and value for its own sake. I do not defend all the views presented by Seneca, and I am treating him as a conversation partner. This is to say, I am not giving an interpretation of Seneca's views on friendship, but making a case for a sort of friendship that can be philosophically justified, which I happen to call "Senecan friendship." It includes some degree of attachment, a shared history, and a common moral and eudaimonic purpose. This is to say, a shared life project. This friendship is loving, emotional, and pedagogical. It does not happen on a day, but requires a constant striving and commitment to progress, not unlike philosophical and ethical projects like the *Letters*.

The question we ask Seneca (and Lucilius) is not only whether we should have friends, but also why. A proper answer needs normative reasons considering what is for human beings; so, Seneca directs us to his *sage*, whose internal and external motions serve as the measure of the human good, nature, and rational choice. First, Seneca will give us a vivid picture of this exceptional human existence, on which he can construct a framework for his theory of goods and value. We shall ask him whether such a complete and self-sufficient life can include love and what place it holds. He will give us two arguments for the friendship of sages, the latter truly completing his account: (i) a pragmatic argument (friendship gives the sage an opportunity for virtuous activity) and (ii) a naturalistic argument (human beings are social beings by nature). The former will serve to reject the Epicurean friendship of utility, while the latter to counter the self-sufficiency of the Cynic sage. Instead, Seneca will defend the natural value of a special *societas*, a companionship for a virtuous life, and its role in true friendship by appealing to analogies with Stoic doctrines of cosmopolitanism. In the last analysis, Senecan

friendship and love do not threaten personal integrity and peace; on the contrary, virtue asks of us that we expand our ethical projects beyond ourselves.

## II. THE SENECAN SAGE AND VALUE

Were we to ask Seneca on this matter, of whether I should love anyone and why, he would respond: “Each time that you want to know what should be pursued and what avoided, look to the highest good (*summum bonum*), the end of your whole life.”<sup>103</sup> So, what is the highest good? For Seneca and the Stoics, it is wisdom, or perfected reason (*perfecta ratio*) “perfect reason is called virtue (*virtus*), which is the same thing as the honorable (*bonestum*).”<sup>104</sup> This gives life its meaning and drives human actions. “How can we be sure of what this is?” Lucilius implicitly asks. Following this request, Seneca delivers this naturalistic proof: by nature, each living being has some distinct quality whose perfection defines its good; the distinct quality of human beings is reason; therefore, perfected reason is the proper good (*proprium bonum*) of the human being.<sup>105</sup> Our human essence contains a concept of what it means to flourish and live well, and the human essence is the rational soul; therefore, human flourishing lies in the development of the rational soul. Already we see the place that naturalism and reason hold in Senecan theory: ethical inquiry is a matter of discovering what the nature of the human soul is and its place in the natural world.

I cannot emphasize enough how comprehensive this “reason” may be. We could take a strict line of interpretation and take it to be restricted to the modern notions of rationality (and perhaps it would be accurate), but we do not need to. A better way to describe this notion might be Seneca’s *bona mens*, literally a good or healthy mind, which includes all mental content in a person: emotions, desires and aversions, goals, values, memories, sense-perceptions, volitions, beliefs, etc. These mental states are not completely internal but have an intentionality, that is, they aim at objects in the world at large. To have a *bona mens* is to cultivate one’s mental contents to be in a proper relation to the world. As such, an intuitively appealing view of this description of happiness is as becoming the best version of myself. If we speak of reason (*ratio*), it is because it expresses that I am intelligent (i.e. I am capable of making meaning of myself and the world) and that all these things in my mind whose development make up this “best version of myself” are intelligible to me as my own. *Ratio*, as the analogous Greek *logos*, implies an account, a narrative. Whenever I say “trying to achieve virtue,” I simply mean pursuing one’s intelligible projects with respect to one’s psychology and world-view; Senecan happiness is to become our “normative self.”<sup>106</sup>

But what content can we give to this “best version of myself”? For Seneca, human flourishing is the life of a sage, who has achieved perfect reason:

“...let’s go back to the principal good and consider what kind of thing it is. The spirit that gazes at what is real, that knows what to pursue and to avoid,

<sup>103</sup> Seneca, *Epistulae Morales* 70.2. In citation, hereafter abbreviate *Ad Lucilium Epistulae Morales* as *Ep* and, unless otherwise stated, provide my own translation of the Latin.

<sup>104</sup> Seneca, *Ep.* 76.10

<sup>105</sup> Seneca, *Ep.* 76.8-9

<sup>106</sup> The “normative self” is the term developed by Tony Long (2007) in his stellar analysis of the Senecan self.



that assigns value to things in accordance with nature and not by opinion—the spirit that injects itself into the cosmos as a whole and casts its contemplation over every action of the universe—the spirit that attends equally to thought and to action—that great and forceful spirit, not vanquished by adversity nor again by the blandishments of prosperity, that does not yield itself up to either but rises above all contingency, all accident.”<sup>107</sup>

We do not need to completely accept this depiction as the best life for ourselves, but we can and must appreciate its place in Seneca’s theory.<sup>108</sup> For him, the sage is intellectually exceptional, for she “knows what to pursue and avoid” and “assigns value to things in accordance with nature.” Therefore, as she is never wrong, the sage is a normative agent, that is, the determinant of human nature and rational choice. Furthermore, her freedom from error grants her freedom from anxiety, “[a] great and forceful spirit, not vanquished by adversity.” This remarkably resilient character speaks to a tranquility and integrity many of us could justifiably desire. As such, even if the Senecan sage is not the sage for us (which she might be) we can still value some of her virtues, namely, her inner peace and strong sense of self.

On this account of the sage and of virtue as the highest good, Seneca builds a theory of value. In the first place, as the highest good, the value of virtue is unconditional; furthermore, the value of all other things is conditional on virtue.<sup>109</sup> In other words, virtue or perfect rationality is the sole source of value in this theory. However, if nothing has value except for perfect rationality, then there is no way of making rational choices. So, a normative agent needs to be able to make value-distinctions. Seneca, always the naturalist ethicist, categorizes value into “primary goods” (things according with nature like pleasure or family), “secondary goods” (things contrary to nature like pain or loss), and “tertiary goods” (neither one nor the other, like picking up a rock).<sup>110</sup> The pursuit of the first and the avoidance of the second are choiceworthy. Seneca calls these the “material of the good” (*materia boni*), for they are the media through which perfect reason becomes actual and in which virtuous activity takes place.<sup>111</sup> After all, without projects to carry out, without things that matter in a human life, without relevant roles to be performed, virtue, or at least its content, is empty. Seneca does not sacrifice our projects for the sake of a consistent ethical theory; they lie at its core. However, we cannot forget that these only have value because of their relation or contact with virtue; therefore, they are always secondary to it. The Stoic sage would be happy in all the materials of the good, even under torture, for she always has what

---

<sup>107</sup> Seneca (trans. Graver and Long), *Ep.* 66.6.

<sup>108</sup> For most of the Hellenistic schools, the figure of the sage lied at the center of their ethical framework. The virtues and way of life of the sage determined what its followers would find valuable and what not. It matters that for the Stoics the sage, the good person, and the happy person all coincide. Furthermore, as Hadot (1995, 345) notes, what distinguishes the Stoic sage, and the Senecan sage especially, from non-sages is her unison with the Stoic God, identified with cosmic Reason and the cause of all things: “The sage of the Stoics knew the same joy as the universal Reason allegorically personified in Zeus, because gods and men have the same reason, perfected in the case of the gods, perfectionable in the case of men, and that precisely the sage has achieved the perfection of reason, making his reason coincide with the divine Reason, his will with the divine will. The virtues of God are not superior to those of the sage.”

<sup>109</sup> Seneca, *Ep.* 119.10.

<sup>110</sup> Seneca, *Ep.* 66.36-37

<sup>111</sup> Seneca, *Ep.* 66.39

she desires most, virtue. But because human nature is constituted in such-and-such way, she might choose to have children and other primary goods and to avoid disease and other secondary goods, if the circumstances allow.

### III. SOME INITIAL PROBLEMS WITH FRIENDSHIP

The challenge for Seneca now is to place friendship in this framework. He needs to show that friendship is a primary good, and, if our intuitions are right, a special one for that matter—not only must it be acceptable to a sage, but also be personally valuable to her. Now, these examples of Achilles, the “anonymous lover,” and myself suggest that our intuitions grasp *attachment* to be a part of friendship, if not a necessary condition. In a Stoic sense, by “attachment” I mean the disposition to engage in whatever behaviors, actions, and emotions root in the belief that “so-and-so is an essential component of my life projects,” that the one I am attached to is, as it were, part of myself.<sup>112</sup> Attachment is a mental state in which I am deeply committed to the idea that my friend holds a mystically high value to me, that the whole structure of my life stands on them, that I could be neither happy nor complete without them. Depending on the circumstances, such belief may become manifest through an unpredictably large range of mental and active states: jealousy, grief, wrath, lust, longing, erratic joy, fear of rejection and loss, obsession. We could picture the anonymous lover being committed to the idea that it would be good and even imperative for her to spend her life with her significant other, or Achilles raging at those heartless monsters who deprived him of Patroclus. In conclusion, to love another person often implies becoming vulnerable.

However, remember what we are striving for: the happy life of the sage. Considering the spectacular tranquility of the sage, we should wonder whether a happy life can include loving another fully, for after all, it does not seem that attachment can coexist with the stellar integrity and self-sufficiency of the sage. Nor does it seem that personal integrity is a non-important project—it is what many of us call today “mental health,” “self-care,” or “sense of self.” So, we must confront a true tension in our value-systems between integrity and love. But, if there is some way of loving a friend that does not entail such loss of integrity, if there is a type of friendship which does not cause those emotional and cognitive upheavals, then such a friendship might be worth rescuing. There is no obvious solution: how can we make sure that we remain whole and complete once we judge that a part of ourselves lies without, that our whole sense of self has been stretched out on the sand to be blown away by the wind? If Seneca can convince us that we can invite others into ourselves without sacrificing ourselves, then he may solve the problems that we have brought to him.

<sup>112</sup> For Seneca, an attachment like that of Achilles, the anonymous lover, or myself would probably resemble an “infirmity” (*morbis*), a persistent judgement such-and-such is more valuable and choiceworthy than it really is (*Ep.* 75.11), and these cause emotion (*adfectus*) which are movements and agitations of the soul, such as delight, anger, grief, desire, etc. In her classic work, Nussbaum ([1994] 1996, 359-401) rightly points out that for the Stoics and Seneca, emotions have some necessary *cognitive* component—each emotion has a certain proposition, which, when assented to by the mind, causes the emotion. It would be categorized as a “passion” to be extirpated and mutually exclusive with a good life if the proposition is a judgement with these properties: (a) it is an attribution of eudaimonic value to some object, (b) it is a false estimation of the value of the object (i.e. I believe that such-and-such is necessary for my happiness when it is not), (c) the object is some external. Sages, who are self-sufficient and have no false beliefs, have no such passions, but could have a limited set of rational emotions in accordance with nature. Arguably, we could imagine that these emotions associated with friendship satisfy these conditions. It is my hope that Seneca may present us with a persuasive defense for an ethical friendship.

## IV. THE LOVE–INTEGRITY DICHOTOMY

In fact, from Seneca’s response, we can infer that Lucilius seems to have similar concerns in letter 9. Seneca introduces the topic thus: “You [Lucilius] want to know whether it is with good reason that in a certain letter of his Epicurus finds fault with those who say that the sage is self-sufficient (*se ipso esse contentum*) and for this reason she does not need (*indigere*) friends.”<sup>113</sup> It makes sense that Lucilius has questions regarding the commitments of Stoicism—after all, he has just read about the special one-ness and companionship of true friendship in letters 3 and 6, yet he is aware of the Stoic stance for the extirpation of the passions. This tension drives the dialectic of the letter, in which Seneca has to distinguish his Stoic sage from Stilpo’s anti-natural “Cynic” sage and his Stoic friendship from the self-interested Epicurean friendship.

Seneca could “solve” this love–integrity dichotomy by either making the self-sufficiency of his sage more flexible, or by abandoning the value of friendship. Instead, he denies the dichotomy and commits to both: “the sage wants to have (*habere vult*) a friend, a neighbor, a person to live with (*contubernalem*), however self-sufficient he may be.”<sup>114</sup> Seneca’s tonal shift from “*indigere*,” implying a lack, to “*habere vult*,” merely implying wish, matters. For Seneca does want to commit to the radical self-sufficiency of the sage who lacks in nothing. She could live well even if she lost her limbs.<sup>115</sup> In fact, by the end of the letter, Seneca drops Stilpo as an interlocutor and turns him into an *exemplum* of self-sufficiency, praising him for asserting that “all my goods are with me.”<sup>116</sup> To make matters worse, he claims such self-sufficiency for the Stoics, and even for the Epicureans.<sup>117</sup>

Nevertheless, we cannot deny the externality of the person whom we love. Actually, I highly doubt that we would love them if they were not somewhat external to us (or do we not love them as free agents independent from our will?). It follows that friendship is not necessary for a good life. But Seneca’s Lucilius points out that an isolated life sounds neither happy nor complete.<sup>118</sup> On the contrary, Seneca responds, the life of a lone sage is like that of Jupiter himself when the universe is ending (and who would call the life of God unhappy and incomplete?). Yet again, Seneca appeals to the harmony between the sage and the cosmos to affirm her strength and self-sufficiency. The comparison brings forth that loneliness does not threaten her happiness but even raises her to the level of God. In this sense, Seneca sides with Stilpo: the sage does not need (*indigere*) friends.

## V. A PRAGMATIC ARGUMENT: FRIENDSHIP GIVES SAGES AN OPPORTUNITY FOR VIRTUE

Does Seneca need to reject friendship for the sake of integrity, then? Perhaps he needs to drop the strong attachment that drives those who grieve inconsolably, but this does not mean that no friendship is possible in Seneca’s ethics. He has

---

<sup>113</sup> Seneca, *Ep.* 9.1.

<sup>114</sup> Seneca, *Ep.* 9.1.

<sup>115</sup> Seneca, *Ep.* 9.4.

<sup>116</sup> Seneca, *Ep.* 9.18

<sup>117</sup> Seneca, *Ep.* 9.19–20

<sup>118</sup> Seneca, *Ep.* 9.16

<sup>119</sup> Seneca, *Ep.* 9.12

a way out: he tells Lucilius that if friendship is choiceworthy in itself, then even a self-sufficient sage would pursue it.<sup>119</sup> But is it choiceworthy? For sure, or at least some forms of friendship are choiceworthy—again the sage wants to have (*habere vult*) a friend, even if she does not need (*indigere*) one. He provides an outline of such friendship by arguing against Stilpo's opponent, Epicurus. In the same letter as he attacks the friend-less philosophers, Epicurus says that a sage would want a friend to help her in illness or imprisonment, making friendship an instrument for pleasure, happiness, well-being.<sup>120</sup> For Seneca, this friendship is one of and for self-interested utility, and so this shall define its beginning and end, "who is brought near because of her utility, she will only be pleasant as long as she is useful."<sup>121</sup> His ironic tone appeals to our intuitions: a union lacking in long-term loyalty and care is neither true friendship nor desirable. In addition, sages, who already have everything that they need, would have no interest in going through this trouble. If there is value in friendship, it cannot lie in its utility. Not even if such utility is eudaimonic, if it contributes to living well, is it desirable. For, Seneca claims, friendship is not desirable to the sage "for the sake of a good life."<sup>122</sup> We agree: yes, our friends do make us happy, but this is not the determinant reason for which the anonymous lover would rationally commit to her partner.

Instead, Seneca proposes that the friendship of sages is true friendship and has rationally justifiable value. He offers a *pragmatic argument* for such friendship: "Even if a sage is self-sufficient, nevertheless she wants to have a friend, if she has no other reason to practice friendship, so that at least such great virtue may not lie neglected."<sup>123</sup> This justifies friendship as a practice of virtue. A sage already has virtue, yet there are certain preconditions needed to actualize it, the material of the good. Again, human life without projects to engage in is a content-less life, and hence *living* itself requires some projects and practices. Virtue injects these, as it were, with goodness, and friendship may be such a project in itself as well as constitutive of other projects. In this way, the sage finds friends valuable in the same way that she finds her hands or even life itself valuable: she needs them to practice virtuous activity, like enduring loss or showing care.<sup>124</sup> This argument seems sound; what makes having roommates worth it, for instance, is the opportunity to make coffee for them every morning.

If the friendship of sages is true friendship, his argument does offer a way out of the love–integrity dichotomy. In fact, Seneca seems to commit himself to the idea that without self-sufficiency there is no true friendship. Elsewhere, he seems fond of the Stoic paradoxes that only sages can be friends and loyal.<sup>125</sup> Now, this argument does place friendship within Seneca's framework of value as a material of the good, for its goodness is expressed in relation to virtue. But to label it a material of the good does not explain friendship's relation to nature and rationality, whether it is choiceworthy and good for human beings. Only if we have such an account of friendship's 'natural qualities' can we understand its unique place in human life, as our intuitions suggest, but also what are the appropriate judgements that we should hold with regards to our friends. Furthermore, note

---

<sup>120</sup> Seneca, *Ep.* 9.8

<sup>121</sup> Seneca, *Ep.* 9.9

<sup>122</sup> Seneca, *Ep.* 9.15

<sup>123</sup> Seneca, *Ep.* 9.8

<sup>124</sup> Seneca, *Ep.* 9.15.

<sup>125</sup> Seneca, *Ep.* 81.12.

that if we follow the argument in letter 66 concluding that all goods are equal to the sage because they stand in equal relation to virtue (i.e. they are all for the expression of virtue), then friendship is just as good for the sage as torture, but torture is not good for human beings. In trying to integrate friendship into a good life, we have made it too thin to grasp. Hence, we need a new argument.

## VI. FOUNDATIONS FOR THE NATURALISTIC ARGUMENT

Since he sees us unconvinced, Seneca complements his pragmatic argument with a *naturalistic argument* to complete his proof of friendship. Before we see his naturalistic defense of friendship, we should understand the place that nature has in Seneca's ethics. We have already seen how his theory of sagehood and value appeals to it. Now that we are trying to grasp his naturalistic argument for friendship, let us see the sort of hierarchy which he builds.

In letter 121, Seneca argues that all living beings since birth have a natural instinct towards self-preservation and an innate perception of its own constitution. This explains animals' evolutionary behavior of sustaining their own life and way of life. From this evolutionary tendency stems an innate love for and attachment to their own constitution (*constitutione sui*).<sup>126</sup> For human beings, this means that we have an instinctual attachment for our rational constitution, and hence, a natural tendency to live as rational beings, whose proper activity is virtuous activity. Though such natural attachment to the self does not imply that it is absolute—for Seneca's sage would commit suicide in some circumstances. Yet, all forms of natural attachments have a reference to this first love of self.<sup>127</sup> If we are to tackle the attachment that we find in friendship, and what sort of a disposition it is, we must do so in terms of this first innate attachment of self. This is to say, a naturalistic defense of friendship will need to appeal to our understanding of who we are as rational animals and to these instinctual impulses.

## VII. NATURALISTIC ARGUMENT: SENECA'S FRIENDSHIP AS COMPANIONSHIP

With this mind, let us return to Seneca's argument in letter 9. As we shall see, as we have a natural attachment to our own self, we have a natural attachment to other people in general. We encounter this towards the end of the letter, where

---

<sup>126</sup>My use of the word "attachment" is purposeful. Seneca uses the passive verb "conciliari" to describe how each animal is bound to its constitution. The word "conciliari" implies that the animal has a friendly disposition to itself as much as it is bound to its own nature. This means that the animal must live as a member of its own species would, literally, according to nature. It is, as it were, that the animal perceives itself as different in concept and yet not distinct in extension (i.e. it perceives a "me" but grasps that "me" as existing within its own body). Note, furthermore, that this "attachment" is quite distinct from the "attachment towards friends" that I spoke of earlier.

When I defined the latter, I referred to it as a cognitive process, namely, as assenting to a certain proposition with respect to the value of the loved one. This attachment towards self seems to be pre-cognitive, namely, a mere impression to which I could assent or not in each particular situation it appears (if this were a cognitive belief that we all hold, the Senecan sage could never commit suicide, but Seneca extensively argues that sometimes the sage would take her own life). Lewis and Short (1879) s.v. "concilio"

<sup>127</sup>Seneca seems to be implicitly alluding to the Stoic theory of oikeiosis here, which accounts for our instinctual impulses towards our self and others. In this view, all animals, including human beings, have a first impulse towards self-preservation. As the human being develops, its constitution is such that it begins to 'appropriate' other people, starting with her family and then country, meaning that she includes them as part of her self-identity. Appropriate and hence ethical action depends on this process. Hierocles's famous 'circles of concern' analogy in Stobaeus 4.671, 7-673, 11 (Long and Sedley 1987, 57G) explains this well: I am encompassed by circles symbolizing how close and how 'mine' something is. My first and closest circle is my own self, including my body. The second circle includes my close family and friends, then outer relations, my fellow citizens, and finally, the entire human race. The sage makes all the circles as close as possible to herself, identifying herself with her roles as friend, lover, and cosmic citizen.

Seneca specifies that friendship is not only as a material of good but also a primary good according with nature:

“It is not her own benefit what brings [the sage] to friendship, but a natural incentive (*irritatio*). Just as other things are agreeable (*dulcedo*) to us innately, so are friendships. In the same way that there is an animosity for solitude (*odium solitudinis*) and a striving for community (*adpetitio societatis*), in the same way that nature bonds one human being to another, there is also an instinct (*stimulus*) which makes us strive for friendships (*nos amicitarum adpetentes faciat*).”<sup>128</sup>

Human nature is constituted to find some things choiceworthy and desirable (*dulcedo*) and to instinctively seek after some things, including community and friendship. Seneca shares this belief with other Stoics and philosophers of his time: a human way of life is a social way of life. Since we hold a natural attachment towards our constitution, this includes our social nature; hence the reason as to why it seems to us as appropriate and good to create and cultivate bonds with other human beings. To suppose that there exists the natural tendency toward friendship would be enough to label it a primary good and hence desirable for the sage. After all, it does not seem empirically untrue that we have a natural instinct towards friendship.

But does this not assume what we need to prove? Indeed, we want to know *why* it is that we become attached to others. For it also seems empirically true that wealth, pleasures, and life are primary goods and that we have a natural instinct towards them, and yet, we *know* that they are not unconditionally good. In fact, some of these items in the list may become unconditionally bad given a morally problematic object, such as the pleasure of being cruel. But such these natural instincts to which we are appealing are object-oriented as well (even the syntax of “the [natural] striving for community” (*adpetitio societatis*) suggests as much); hence, we need more to make sure that this is not similar to the case of the pleasure of cruelty. It matters then that we have a proper account of friendship. Otherwise, the possibility exists that all love is deception and based on false belief.

At the same time, there is a practical/psychological concern. We might not realize it, but in our everyday lives, we *do* analyze what these “instincts” are, what they mean, and what they aim for. If we did not, then we would not feel the urge to be in community and in friendship. For the attachment and emotions that make up friendship do require a certain assent, a commitment, to the belief in the importance of the friend. We need good reasons for this (mostly cognitive) assent. Yet, these natural impulses themselves do not provide good reasons for the assent, even if they allow for the assent by providing the proposition to be assented. They are mere impressions on our minds begging us to analyze them so that we can say ‘yes, it is true *that* I should be with other people and avoid being alone.’ Or are all friendships acceptable and ‘natural’ and require no justification of why they are so? If so, how are we to respond to our challengers, or to the anti-natural Stilpo and his objections against the value of friendships? ‘Obviously, we want friends’ is not an adequate answer, even if it is true.

---

<sup>128</sup>Seneca, *Ep.* 9.17

Nevertheless, I would argue that Seneca's naturalistic argument has some deeper strengths. What we see in his naturalistic argument is that the essence of Senecan friendship is *societas* (companionship or partnership). I shall use the Latin word because, more so than the English one, it implies a functional concept—it is a partnership that comes about for the sake of a specific purpose.<sup>129</sup> This *societas* is the object of our impulse; so, we may better grasp the strength of Seneca's naturalistic defense of friendship by further analyzing his concept of *societas*.

At this point, Seneca has pointed to a two-fold social nature: he invites us to distinguish between a natural instinct towards what I shall call 'general *societas*' and that towards 'particular *societas*.' The difference between general and particular is the difference between the "*stimulus*" for "animosity for solitude and striving for community" and that which "makes us strive for friendships" referenced in *Ep.* 9.17, between the bonds that may hold among fellow humans and countrymen and those among friends and lovers. The psychological motivational force of general *societas* could be a sense of loyalty for the community and nationalism, or fellow-feeling, sympathy, and a sense of humanity, while that of particular *societas* is also attachment (and all the emotions this implies such as empathetic delight or grief) and love, not just for humanity, but for an individuated and irreplaceable person.

#### VIII. NATURALISTIC ARGUMENT: EPICUREAN SELF-INTEREST, THE SECOND SELF, AND COSMOPOLITANISM

So, to complete this defense of friendship, we need to establish two things: in a first place, that if general *societas* is natural and choiceworthy, then particular *societas* is as well, and furthermore, a proof of the value of general *societas*. To do so, let us turn to a different letter, 48. The letter's introduction suggests that Lucilius has asked a favor from Seneca, namely, that he solve a logic puzzle for him. Seneca responds that he needs some time to figure it out because of its difficulty; but more importantly, Seneca seems reluctant to engage with Lucilius in useless logic puzzles, for he is more interested in discussing ethics, not logic as Lucilius does; and so, he says "and while one thing would be beneficial to you (*expediat tibi*), another would be beneficial to me" to which he wittingly adds: "am I talking like an Epicurean again?"<sup>130</sup> In this comedic and somewhat ironic exchange, Seneca returns to his strawman depiction of Epicureanism from letter 9 as mere utilitarianism. For Seneca, Epicurean practical rationality is a matter of bringing about the greatest benefit to oneself. Epicurean friendship, then, exists for the sake of personal well-being and satisfying self-interest. But for us, the utilitarianism of Epicurus may not mean that altruism, justice, and virtue in general are not possible in this model. Still, our intuitions suggest that there might be a problem if these 'so-selfless' moral concepts have become a mere function of my own pleasure. In fact, the utilitarian-Epicurean adversary would respond that it is wrong to say that Epicurean friendship is necessarily neither profound nor deep. They could point to the fact that there is some utility in cultivating a long-term, solid, and deep friendship. They would even add that this is how *all* friendships come about, as functions of self-interest and well-being.

<sup>129</sup>Lewis and Short (1879) s.v. "*societas*."

<sup>130</sup>Seneca, *Ep.* 48.2.

We have already explored the first of Seneca's responses: friendships whose origin lies in utility ends if breaking up would benefit one of the friends.<sup>151</sup> But again, the Epicurean-utilitarian responds, the antecedent is never satisfied for healthy relationships for whom there is no benefit in breaking up—we see this in that it is not impossible to conceptualize long-term, solid, profound friendships whose purpose is the well-being of each individual in the relationship. Unlike Seneca, I do grant this rebuttal; yet, I see something deeply ethically disturbing in this model. I shall propose that we should not think that this is the only possible model of friendship.

We return to Seneca's letter 48, where we can see that truly Senecan friendship operates with a radically different conceptual model, one that shifts the practical rationality behind friendship from individual utility to the fulfillment of common life projects, one that even advocates for a sort of dissolution of the individual self. After his rhetorical question bringing up Epicurean utilitarianism discussed earlier, Seneca adjusts his comment:

“In reality, the same thing that benefits (*expedit*) you also benefits me; I am not your friend, if your affairs are not also mine. Friendship establishes a fellowship (*consortium*) in all things between us. Nor is there good or bad fortune for the individual; we live in common (*in commune vivitur*). Nor can anyone live happily who thinks of himself only, who turns all things to his own advantage; you must live for another, if you want to live for yourself.”<sup>152</sup>

In so far as Lucilius and Seneca are distinct individuals with proper desires, they have divergent interests (i.e. discussing logic vs. ethics, respectively). Once this is pointed out, Seneca makes a point of correcting himself, for, in so far as they are bound together by their friendship, all their interests converge. What justifies this shift? In contrast to the Epicurean mode of ethical thinking and friendship, Seneca portrays his friendship to Lucilius as an *expansion of his own self*. The shared way of life, that friends live in common, implies that whichever natural instinct to love myself I have, this same impulse manifests itself in a natural instinct to love my friends, those persons whom I have made my companions and with whom I share my life projects. Here we see the essence of Senecan friendship: it is a *societas* in which the friends include each other's projects and even identity in their own. At a practical level, there is a notion of a common good, of individual sacrifice regardless of any benefit—this shift in practical rationality is what drives me to move my friend's queen-size bed across town independently of whether he will repay me or whether it brings present or future pleasure; it is what makes the anonymous lover fear the loss of her beloved as if she dreaded the thought of a part of herself being robbed from her.

Nevertheless, if it really is an expansion of the self onto another person, how is this sense of a *societas* and this natural tendency to bring about the common good of the friends different from Epicurean friendship? Isn't the Senecan friend merely turning her friend's advantage into her own such that she always acts in her own “self-interest”? Isn't this utilitarian practical rationality as well? The Epicurean

---

<sup>151</sup>Seneca, *Ep.* 9.9.

<sup>152</sup>Seneca, *Ep.* 48.3.



could turn Seneca's argument against him, responding that if a person were to abandon her friend when it is "useful," as Seneca worries that Epicureans do, then she is just foolish and does not know what "true utility" is and what is to her own advantage. Epicurean sages know that the interest of the friend is the interest of the individual, hence why they are capable of forming long-lasting friendships.

Still, it seems to me that there is a radical distinction in the way that the Senecan model invites us to experience our own self and its place in the world. Insofar as the Epicurean good is pleasure and, to the extent that Epicurean friendship is hedonistic, it is self-interested and hence problematic. In contrast, Seneca wants us to go beyond ourselves, to transcend our body, soul, and individuated self.<sup>133</sup> For the Senecan good is perfected reason, which demands a deliverance of the self to the whole human race and the cosmos. It is remarkable that in this passage Seneca immediately moves his attention to the relation between love of friends to love of humanity: he continues this expansion of the self to the other to include the whole human race, moving from the particular *societas* of friends to the 'universal *societas*' of humankind. He adds:

"This sense of companionship (*societas*) which joins human beings together and holds that there is some common law of the human race (*commune ius generis humani*), when it is cared for sacredly and attentively, also greatly contributes to the cultivating of that companionship of which I was speaking, which is internal to friendship. For he who has much in common with a human being (*homine*) will have everything in common with a friend (*amico*)."<sup>134</sup>

Note the shift from the more specific second-person of the earlier passage addressing Lucilius to the third-person of this one, referring to an impersonal cultivator of the *societas* of the human community. Without doubt, this must be the sage. By putting his concept of friendship in relation with his concept of the human race, Seneca distinguishes his theory from the Epicurean one. Indeed, it could be, as the Epicurean-utilitarian adversary proposed, that if we love our friends because they are life companions, second selves, then Senecan

---

<sup>133</sup> It is beyond the scope of this paper to address all concerns regarding the Senecan self. For now, I shall only take a stance on the disagreement between the Foucauldian and Hadotian interpretation of the Senecan self in the latter's "*Réflexions sur la notion de culture de soi*." In the Foucauldian model, the Stoic method of Seneca seems to cultivate the self for the sake of reflexive pleasure and happiness, even if an elevated one (what Seneca calls *gaudium* in letter 24). In this model, virtue, friendship, and philosophy (of virtue and friendship) aims at a pleasurable experience of the self, and for that matter, a self that seems to be ontologically distinct from the external world and its community and hence ethically self-concerned. On the other hand, Hadot argues in a first place that Senecan joy (*gaudium*) is fundamentally distinct from pleasure (*voluptas*). He points out that for the Stoics, and for Seneca in particular, self-cultivation is not a matter of being aware of myself and improving myself for the sake of a happy existence with myself. It is a matter of seeing myself in relation to the larger cosmos, as a rational part belonging to a rational Whole. The object of Senecan joy is this relation, not "Seneca," but a "transcendent Seneca" connected to the larger human community and intelligent universe. Hence, Hadot's "transcendent self" is not concerned with itself as a distinct entity that requires attention for its own sake *per se*, but with itself as a part of the community, and hence has an outward ethical orientation. I believe that the same problem manifests in Seneca's defense of friendship. The cultivation of friendship is not a part of the project of the cultivation of the self, but the way in which we dissolve the borders of the individuated self and become part of a grand city and narrative, a cultivation of humanity. The object of Senecan friendly love is a transformed self which includes a larger world than its own soul and body and sees itself as identical to a cosmos. Hence why Seneca seems to connect it to the way that we relate with the larger human community.

<sup>134</sup> Seneca, *Ep.* 48.3.

friendship does not differ from Epicurean-utilitarian friendship. ‘You simply consider your friends’ well-being to be part of your overall-well-being, such that pursuing a common project with them brings *you* pleasure, too,’ they say, ‘you are simply including more preferences, those of your friend, under the set of your personal preferences.’ This could be true of imperfect, non-normative friendships among non-sages like Achilles, Seneca, or myself. But for sages, the impersonal cultivators of the universal *societas*, friendship seems to be a manifestation of her communion with the entire cosmos and human race. Friendship is the beginning of the expansion of the self, not as a dissolution of her own self, but as fading of the borders that delineate her mind and body and, as they wilt, they begin to include more beings in her spheres of concern.

#### IX. NATURALISTIC ARGUMENT: THE ANALOGY OF FRIENDSHIP AND COSMOPOLITANISM

Note how our conceptual understanding of the categories of ‘societas’ has changed: earlier when we close-read 9.17, Seneca claimed that we have a natural instinct towards *societas* in general, namely, to live with other human beings; now in 48.4, we see that this is not just an aversion towards isolation (*odium solitudinis*), as if we could satisfy this instinct by merely living with some and shunning others (i.e. tribalism and nationalism), but even that this instinct is for a *universal societas* with humanity as a whole, resulting in certain moral behaviors, such as humane treatment of all people, and certain emotions of fellow-feeling, whose object is humankind as a whole or the humanity in specific persons. Yet, while the movement of Seneca’s writing and of Stoic moral development suggests that we arrive at this *cosmopolitanism* (i.e. sense of global human community) through friendship, Seneca’s actual words indicate that it is through the cultivation of the former that we cultivate the latter. So, we need to further explore the extent to which this relation holds to draw any conclusions. I shall explain three distinct ways in which this relation between friendship and cosmopolitanism appears: first, an epistemological relation (that we grasp one through the other), secondly, a metaphysical relation (that they are the sort of thing and have the same nature), and finally, a moral relation (that the conduct and emotions involved are identical and hence the ethical justification of each is the same).

It is fitting that we begin with the epistemological relation, for it sets friendship and cosmopolitanism as *analogous concepts* and gives us a framework to understand any relation between them. What we need to understand is how Senecan moral knowledge can be constructed in the first place. In letter 120, Seneca explains that we build our understanding of the good and of virtue through the critical observation of historical and literary *exempla*, by comparing and contrasting their merits and failures.<sup>135</sup> We conceptualize ‘virtue’ and ‘what is good

---

<sup>135</sup> It is beyond the scope of this paper to provide an account of Seneca’s theory of our conceptualization of the good and the complexities of how this cognitive process works in Stoicism. It will suffice to point out that moral knowledge comes about through abstraction from “pre-conceptualized impressions” of virtue, including analogous thinking, as Inwood (2005, 299) puts it: “Seneca thinks that we can explain how reflection on the experience of imperfect but laudable agents can veridically generate a notion of perfect goodness. We do this in part by careful abstraction from their acts, aided by a providentially natural tendency to focus on the good. But more important is the role of an ideal of a perfect agent (even one un-instantiated or attested only in literary tradition) in focusing our attention and helping to distinguish those features which constitute goodness.”

for human beings' by observing patterns of behavior, discerning praiseworthy and blameworthy figures, like Cato or Socrates, and inferring some ethical conclusion. This is a complex, difficult process, one that seems to span across the history of civilization. Fortunately, our mental apparatus includes some cognitive tools to achieve this goal, including *analogous thinking*. Indeed, Senecan moral knowledge comes about through analogies, for example, "we knew that there was health of the body; from this we cognized that there is some health of the mind."<sup>136</sup> How is this useful for our argument? In our inquiry, we have already established an analogy between friendship and cosmopolitanism which lies in the concept of *societas*: as business partners come together in a *societas* for their business projects, friends come together in a *societas* through which they have a shared life project; as there is a sense of *societas* among friends, there is a sense of a universal *societas* among human beings. But more interestingly, Seneca comments that "we" (i.e. human civilization) *learned* the virtue of justice, of giving each their due, by noting the man who was "beneficent (*benignum*) towards his friends."<sup>137</sup> In other words, our concepts of 'humane treatment,' of 'our obligations to human beings,' of 'the bond which holds the universal *societas* of humankind' is built on our understanding of the friendly behavior of sages, on the concepts of 'ideal friendly treatment,' of 'ideal obligations to friends,' of 'the bond which holds the particular *societas* of friends.'

Moreover, I suggest that the reverse relation stands as well: knowledge of this cosmopolitanism results in knowledge of friendship. We will use this approach in this part of the conversation: I hypothesize that a theory of cosmopolitanism can yield some proper account of friendship. After all, our conversation and inquiry seek to establish some knowledge of friendship, namely, what ethical justification we can provide of it, its moral nature, its cause, its value in human life. If we believe Seneca in 48.4, knowledge of "having much in common with any human being" may become knowledge of "having much in common with a friend." So, what place does this have in our argument? Since the value of friendship is to be found in the normative friendship of sages, then it is key for our inquiry that we grasp the sageful love; in addition, since we grasp sageful love through its analogy with the sense of cosmopolitanism of the sage, we need an account of the latter.

## X. NATURALISTIC ARGUMENT:

### THE ESSENCE OF FRIENDSHIP AND COSMOPOLITANISM

This moves us to my metaphysical interpretation–reflection of what the universal *societas* actually is shall yield an argument for its value and, in addition, the value of friendship. Now, in 48.4, Seneca points us to the *societas* of the human community to understand the *societas* of friendship. What is the foundation of the former *societas*? Seneca responds: the "common law of the human race" (*commune ius generis humani*). For Seneca, and for Stoics in general, this is a philosophically charged term alluding to Stoic cosmopolitanism—not only a mere sense of a human community, but a physics concept in its own right. Seneca's use

---

<sup>136</sup> Seneca, *Ep.* 120.5.

<sup>137</sup> Seneca, *Ep.* 120.10.

of that term should then imply that his understanding of *societas* builds on Stoic physics.

Let us then give a brief but sufficient overview. In the Stoic and Senecan physical world-view, the universe is made of matter and reason, the latter being cosmic Reason, the Stoic God, which shapes and directs an ordered cosmos as if it were its ruler.<sup>138</sup> Such Reason manifests itself in each human being through the form of our rational soul; hence, all human beings are of the same kinship and of divine origin.<sup>139</sup> We are all divine beings. Slaves and slave-owners, men and women, Romans and Carthaginians, share in such a cosmic community insofar as they all possess a human soul which can grasp its place in the cosmic causal chain and intelligent universe. In this way, the laws of physics *are* the common laws of the human race.

But if you and I do not believe in this cosmology, we might prefer this further argument for the existence of a human community. The Senecan sage, aware of our teleological human nature, recognizes that all human beings have a common purpose, and hence a reason to come together in a *societas*. For Seneca's rational soul is a developable capacity with a specific aim, its own cultivation and perfection, virtue. Or, if nothing else, we have a common purpose in a happy life, for who does not want what they want? In this way, our natural constitution for sociability lies in our purposive human essence, a rational soul, or as a minimum, a soul seeking a happy life. It follows that the final cause of the universal *societas* is such a shared end, for we (should) mutually recognize each other as happiness-seeking beings whose flourishing, even if plural, always refers to our human soul begging for development, which we call virtue. Normatively speaking, the cultivation of virtue binds humanity together. Hence, "the first promise of philosophy is fellow feeling (*sensum communem*), a sense of togetherness among human beings."<sup>140</sup>

In so far as it is true that there is some world city, and in so far sages have perfect reason, the sage grasps this cosmic panorama, for, to reiterate, her soul "injects itself into the cosmos as a whole (*toti se inserens mundo*)."<sup>141</sup> As her perfect reason is as it were one with the cosmic one, so it is one with all its particular manifestations in human beings as a whole and in each human being. Hence, she perceives herself to be linked to all human beings. It matters that she grasps all these facts about herself and others. For it is this recognition which creates that sense of a universal *societas*, of a fellowship among all people. It should come as no surprise, then, that she regards the whole world (*mundus*) as her republic.<sup>142</sup> From her harmony with rationality grows forth her fellow-feeling, and from this "theory of everything," an ethical theorem: the sage's sense of universal *societas*.

Our conversation with Seneca has yielded that universal *societas* is a recognition of a common purpose among the human race, one which the sage is very much right to feel and cultivate. Even if we do not share Seneca's cosmology or even teleology, this is a meaningful concept, if nothing else, for its

<sup>138</sup> Seneca, *Ep.* 65.2,25; note that the term "cosmopolitanism" does aim at creating an analogy between the universe and a political structure, though for our purposes it may not need to be more than an analogy. Though it could serve as the basis for an argument for a world-state, or an internationalist politics, Stoic, and Senecan, cosmopolitanism in this context refers to this understanding of the cosmos as directed by some intelligent ruler. It does, as political structures do, imply a sense of community of all the citizens of the cosmic city, including all human beings.

<sup>139</sup> Seneca, *Ep.* 44.1.

<sup>140</sup> Seneca (trans. Graver and Long), *Ep.* 5.4.

<sup>141</sup> Seneca, *Ep.* 66.6.

<sup>142</sup> Seneca, *Ep.* 68.1.

illuminating value. From this, we can see that the particular *societas* embedded in the friendship of sages is *the same sort of thing*, a common life purpose, a shared life project. Like the sage recognizes a common goal with humanity, so will she recognize a shared goal with another sage. Namely, this common goal is a certain way of life, one in which they may practice virtuous activity together. In this model, the value of the sageful friend lies in her role as fellow-sage, as somebody to encourage the other to realize her potential: “The sage finds the activation of her virtues valuable (*opus est*): and in the same way that she makes herself active, she is moved by some other sage.”<sup>143</sup> This answers the question of what benefit can friendship bring to a life that is already perfect: Seneca explains to Lucilius that sages provide each other for opportunities to use their virtue, share their knowledge, and delight in the good of the other.<sup>144</sup> This virtuous life project binds them together in the *societas* so-essential to friendship. In this sense, friendship of sages is a qualified form of the fellowship that exists among human beings—at its most essential level, particular *societas* truly is a perfected universal *societas*. For the sense of companionship within the human community is cared for through mutual recognition of rationality, likewise companionship between sages through mutual recognition of perfect rationality: “therefore just as reason is needed to activate (*movere*) reason, to activate perfect reason, perfect reason is needed.”<sup>145</sup>

If the epistemological interpretation of the relation between cosmopolitanism and friendship gave us a tool for inquiry in the analogy between them (that they are like each other), this metaphysical one above goes a step further and establishes a closer identity. Both are, at their core, a teleological union of human beings; it matters that we can assess and somehow embrace the truth-value of this proposition: ‘there exists some common end for *us*’—and not just any end, but *the* end of human life. We have not strayed from Seneca’s advice: “each time that you want to know what should be pursued and what avoided, look to the highest good (*summum bonum*), the end of your whole life.”<sup>146</sup> And so, the sage says to herself: ‘*this* person, *that* person, *all* people, and myself have a common aim in our human flourishing,’ or to her sageful friend, ‘you and I have a shared life project.’ It matters that we have found expressible concepts of commonality, for the sense of cosmopolitanism and of friendship to develop, we need a certain recognition on her part. For, when these propositions are sensed and cognized, they grow into the family of emotions and behaviors making up the sage’s sense of cosmopolitanism and friendship.<sup>147</sup> From this emerges the moral relation between these two—in closer inspection, there seems to be a strong resemblance between the conduct and affection that makes up the practice of cosmopolitan citizenship and of friendship.

---

<sup>143</sup> Seneca, *Ep.* 109.3.

<sup>144</sup> Seneca, *Ep.* 109.1-5.

<sup>145</sup> *ibid.*

<sup>146</sup> Seneca, *Ep.* 70.2.

<sup>147</sup> Stoic theory of emotions and action is a cognitivism; namely, that they are the result of those propositional beliefs that we assent to. As Graver (2002, 91) rightly argues, emotions and behaviors come about from practical syllogisms whose premises need to be perceived and cognitively assessed as true at some conscious or unconscious level. Note, for example, how this practical syllogism results in grief: ‘the death of my child is an evil for me;’ ‘when something which is an evil for me has just occurred, it is appropriate for me to feel mental pain;’ ‘my child has just died;’ therefore, ‘it is appropriate for me to feel mental pain now.’

## XI. NATURALISTIC ARGUMENT: THE VIRTUES AND EMOTIONS OF FRIENDSHIP AND COSMOPOLITANISM

I admit that I cannot give a full account of what cosmopolitan citizenship entails here; in fact, my approach of explaining friendship through it might seem counter-intuitive, for we more naturally understand cosmopolitanism based on closer and more familial bonds, such as patriotism or family and friendly love.<sup>148</sup> I hope, nevertheless, that it may suffice to assert that:

(a) the sage who possesses all the virtues conceivable to human beings—in particular the Senecan virtues of moderation, courage, wisdom, and justice—does in fact see herself as having something in common with any and all persons,

(b) that from such recognition of commonality and virtues springs forth a certain, as it were, sense of civic duty and proper behavior towards strangers in so far as they are happiness-seeking creatures to whom we may show beneficence,

(c) that such conduct constitutes what we may call humane treatment.

Moreover, I argue that the sage goes beyond the mere actions of the human rights activist or warrior of justice: she quite literally engages in ‘philanthropy,’ the love of humanity. When she is walking down a street in a country far from hers and makes eye contact with another person, she can see something familiar; she then recognizes their conjoined participation in the cosmos as human beings and experiences a variety of ethically positive emotions, such as a special goodwill for the stranger.<sup>149</sup>

This phenomenon we might identify with ‘fellow-feeling.’ It would, depending on what the circumstances demand, manifest in a moral sympathy that drives the sage to beneficence (if her aid is needed), clemency (if her moderation is needed), fortitude (if her courage is needed), etc. If we want to take it further than Seneca and the Stoics, we might include compassion for those who need our help and indignation at the injustices of the world.<sup>150</sup> What matters is that we see in the sage, who, to reiterate, is our normative agent, a certain affective and behavioral virtuous disposition towards human beings and even strangers; that in fact, in the absence of the thick ties of blood or tribe, this moral link might be the only way in which the sage relates to strangers, as citizens of the cosmos. It seems then that to cultivate our disposition to feel these cosmopolitan emotions, and to

---

<sup>148</sup> It should suffice to note that we might ground this moral conduct on a sense of human rights and respect for human beings regardless of race, ethnicity, gender, sexual orientation, religion, or nationality. Nussbaum (2019, 64) has pointed out that there is a strong resemblance between this approach found in the Cynics and Stoics and modern philosophers on universal duties such as Kant, for, she claims, “Cynic/Stoic cosmopolitanism urges us to recognize the equal, and unconditional, worth of all human beings, a worth that the Stoics grounded in (practical) reason and moral capacity.”

<sup>149</sup> Although the Stoics argue for the extirpation of the passions, overwhelming affects with false propositional content and whose object is an external good, they believe that the sage experiences *eupathe*, or ‘good affects,’ rational feelings whose object is truly worth feeling for, including virtuous activity, cosmic harmony, or the good for. Among this is Seneca’s ‘real joy’ (*vera gaudium*) at one’s own virtuous self. As Graver (2007, 57-65) has argued, ‘other ‘eupathic emotions’ include: goodwill (the lingering wish that good things happen to a person), good spirits (the joy at the management of the universe), or moral shame (caution against correct censure).

<sup>150</sup> Nussbaum (2001, 304-305) points out that Stoics explicitly consider pity to be ethically problematic passions to be extirpated. Stoics claim that the cognitive structure of this passion to be incompatible with the perfect reason of the sage. Pity involves a false evaluation, namely, a recognition that the pitied person’s suffering is due to externals, that it is justified, and that it could happen to me. Since the sage only considers moral evils as genuine, she arguably does not experience this sort of pity for those who suffer in the world, whether because of political tyranny, economic inequality, or social ills. Now, it is clear that the Senecan sage would experience a form of eupathic fellow-feeling and moral sympathy; nevertheless, I admit that it is a non-trivial question whether she would feel pity or anger the way we think of it. Though it does strike me as possible and genuine, it is beyond the scope of this paper to argue for a eupathic compassion.

feel them as the sage would, is to cultivate that universal *societas* with Seneca's impersonal "human being" (*homo*) in the 48.3-4 passage.

And is this not similar to the way that we cultivate the particular *societas* that I have with a friend? Seneca's tone later in 48.4 praises her for whom "a human being is as a friend" and condemns her for whom her "friend is not a human being." If our intuitions are right, the cosmopolitan citizens' 'eupathic emotions,' or rationally justifiable and ethically valuable feelings proper of good human beings, are not too far from that of the sageful friend. The virtues of friendship and of the cosmopolitan citizen seem too similar (they strike me as particular manifestations of the Stoic virtues). They both, after all, come about from a recognition of certain propositions about rational *societates*, from the realization that there truly is a shared project among friends and among the human species.

It seems, as Seneca told us a while ago the first time we read Ep. 48.3, that "the person who has much in common with a human being (*homine*) will have all things in common with a friend (*amico*)." How so? Seneca is not only making an *a fortiori* argument, but he is bringing forth a necessary component of his model of friendship. For true friendship is not Epicurean-utilitarian, based on self-interest, but it is a particular *societas* between two self-sufficient persons who, paradoxically, wish to expand their own selves. If this companionship requires that sense of cosmopolitanism, it is because the latter is a cultivation of my affective and behavioral dispositions towards human beings, requiring shifts in my beliefs about my own self and my place in the world. In particular, these beliefs speak to a spiritual others' understanding of the human soul as a part of a larger Whole to which we all belong, truly a cosmic city. This shift in viewpoint is what justifies the expansion of the self which consolidates our Senecan attitude that friends live in common. Hence, the Senecan naturalist defense of friendship: friendship is valuable to the extent that it is a sense of companionship in a virtuous life, and this companionship has a special place in the structure of the life of a person in tune with the cosmic Whole to which she belongs. For a virtuous agent lives transcending herself, and friendship is a constitutive part of that project.

## XII. THE LOVE-TEGRITY DICHOTOMY, REVISITED

We return to the love-integrity dichotomy which separates Epicurus and Stilpo. Seneca promised us that this would be shown to be false. He has fulfilled his promise, albeit in a peculiar manner. For in the first place, this picture of the sage as self-sufficient has been questioned. The sage is not self-sufficient *per se*, but an actor and contemplator of the affairs of a self-sufficient world. I reiterate that we need not, unless we wished otherwise, to accept Stoic cosmology—all we need to do for the argument to follow is to accept that we are part of something larger than ourselves, even to merely embrace ourselves as beings of a species. This need not be a spiritual belief (in fact, the Stoics were physicalists themselves); and especially in today's world, connected by international law and a global pandemic, where we have an understanding of the shared humanity of all people and human rights, of our moral actions in the world, of the role of psychology and sociology in our self-cultivation and ethics, any claims to solipsism lose any solid intuitive justification.

And once self-sufficiency does not mean radical individualism, solipsism, or any other '-isms' that isolate the soul, to invite others into ourselves does not

threaten the concept of personal integrity. Nor does the Senecan friendship compromise our ethical intuitions, as the Epicurean-utilitarian one; on the contrary, it exists for the sake of the project of virtue. In the last analysis, being good demands an expansion of the self. Seneca has brought us back to friendship as a primary good, something which allows the sage to actualize her virtue and is according to nature, but now ‘according to nature’ has a new meaning. A person who lives following nature prefers to have good friends, for it is an expression of the purpose we all share because of our human essence. Hence, Seneca ends up taking a middle stance between his Epicurean and Cynic interlocutors in letter 9: Senecan friendship differs from the Epicurean friendship because its foundation is not self-interested utility but in affective altruism among the self-sufficient; at the same time, Seneca has moved away from Stilpo’s Cynic sage as well, for although he agrees that the sage does not need a friend to live well, friendship is desirable to her by nature. Seneca tells us in *Ep.* 9.3 that Stilpo fails to account for natural affects, and so he also fails to present human beings as naturally social and loving. In contrast, Seneca’s portrayal of *who we are* strikes a chord of compassion, care, and humanity.

Where has the argument led us? Seneca has taken us on a walk through the cosmos to show us the place of friendship in human life. Sages do love and come together in friendship for the sake of sharing their honorable life projects, for human beings were born to live together. Still, they do not feel the passionate attachment of Achilles or myself, but instead a profound appreciation of the friend. Is true friendship not for morally normal people then? Perhaps not; and yet, for those of us striving for goodness and happiness, like Seneca and Lucilius, friendship is essential, even necessary. Through friendship we cultivate ourselves and realize our ethical projects, learn how to love humankind, expand ourselves to another and, through the other, to the world. We craft ourselves with friends; after all, Lucilius is such a second self for Seneca in his *Letters*. Its genre provides on its own an account of the value of friendship of philosophers, for his letters are essays whose purpose is the ethical growth of two friends. Such a friendship is not yet complete, for they are still “learning to love” each other.<sup>151</sup> Nor is it static or lacking in conflict; still, it runs deep. Finally, even if Seneca’s arguments did not satisfy us, his style does something which no argument by itself can. He animates his philosophical arguments with this literary dialogue of friends, and as we observe them unfold, this we can conclude: in the *Letters on Ethics to Lucilius*, philosophy, without which “no one can live a good, or even tolerable, life,” appears as a shared activity of friends, and friendship as a philosophical and ethical project.<sup>152</sup> We could ask for no better defense of love, friendship and companionship, and a shared life from philosophy.

---

<sup>151</sup> Seneca, *Ep.* 35.1-3

<sup>152</sup> Seneca, *Ep.* 16.1.



## BIBLIOGRAPHY

Graver, M. *Cicero on the Emotions: Tusculan Disputations 3 and 4*. Chicago: University of Chicago Press, 2002.

———. “The Pathetic Syllogism.” In *Stoicism and Emotion*, 57-65. Chicago: University of Chicago Press, 2007.

Hadot, Pierre., *Qu'est-ce que la philosophie antique?* Barcelona: Folio/Essais, 2018. First published 1995 by *Éditions Gallimard*.

———. “Réflexions sur la notion de ‘culture de soi’.” In Michel Foucault *Philosophe* 261-270. Paris: *Éditions du Seuil*, 1989.

Inwood, Brad. “Getting to Goodness.” In *Reading Seneca: Stoic Philosophy at Rome*, 271-301. Oxford: Clarendon Press, 2005.

Long, A.A. “Seneca on the self: why now?” In *From Epicurus to Epictetus*, 360-376. Oxford: Oxford University Press, 2006.

——— and David Sedley. *The Hellenistic Philosophers*, vol. 1. Cambridge: Cambridge University Press, 1987.

Nussbaum, Martha C. *The Cosmopolitan Tradition: A noble but flawed ideal*,. Cambridge: The Belknap Press of Harvard University Press, 2019.

———. “The Extirpation of the Passions.” In *The Therapy of Desire: Theory and practice in Hellenistic Ethics*, 359-401. Reprinted 1996. Princeton: Princeton University Press 1994. Page references are to the 1996 edition.

———. *Upheavals of Thought: The intelligence of the emotions*. Cambridge: Cambridge University Press, 2001.



---

---

The Limits of Artistic  
Autonomy: The Case of  
Modernist Public Art

---

---

*Alice Pessoa de Barros*  
*McGill University*

## FOREWORD

*Editors Weston Barker, Liam Galey*

In 1917, the artist Marcel Duchamp took a factory-made urinal, signed it ‘R. Mutt’ and then submitted it to an exhibition. Years later, we remember the urinal, titled ‘Fontaine’, as a pivotal piece of avant-garde art, one which had a meteoric impact on the artistic establishment and is perhaps the quintessential example of an artist being entirely free in the production of her work.

Notably, Duchamp’s work was in a museum, and was, at least in its original iteration, never installed on a public street corner or crosswalk. Perhaps it was better suited as a non-public work of art, as if left unattended, it might have had its functional appearance tested. However, suppose it was displayed in public - should the work have been granted the same consideration it was granted on museum display? Further, does an artist have the same freedom, or ‘artistic autonomy’, in the creation of works of public art as they do in works meant for private display?

In ‘The Limits of Artistic Autonomy: The Case of Modernist Public Art’, Alice Pessoa de Barros critically examines the idea of the avant-garde artist as free in the production of her works. Positing the private exhibition space as one which can entertain challenges to the artistic status quo, Pessoa de Barros proceeds to examine the public space, establishing it as one which necessarily hinders artistic autonomy. Further discussing specific genres of public art, Pessoa de Barros aptly highlights the narrowing of artistic autonomy, as the standards for public exhibition are complemented by the standards of a genre therein.

In addition to offering intriguing examples of public and private works of art for the reader to consider and digest, Pessoa de Barros engages with questions with considerable gravity for the artist and art-consumer alike. Navigating various definitions of art and freedom, weighing the purpose of creation with the purpose of display, the paper achieves a sublimity in its argumentation while remaining accessible to those without prior knowledge in aesthetics. As the paper could function as a field-guide for the avant-garde artist considering submitting a work for display, so too it could be an introduction to increasingly germane questions as art becomes more rooted in controversy.

Duchamp’s ‘Fontaine’ and the avant-garde are not dead by Pessoa de Barros’ judgement, but they must be properly contextualized when entering into the domain of public art. Perhaps there will rarely be the need for such contextualization or conflict, as “if an avant-gardist work is broadly enjoyed by the public, it becomes an unsuccessful avant-gardist work”.

## ABSTRACT

This paper will look at some of the limits surrounding the artistic freedom granted to modernist public artworks by contesting an argument raised by Caroline Levine in her paper “The Paradox of Public Art”, which we will call the “artistic autonomy” argument. She uses this argument to show that art is not for the people but for the artist and their creative autonomy. I will first argue that protecting the artist’s creative autonomy does not grant them immunity in the case of public works like it does in private settings, because it can justify putting any work in public and because a work of art comes with a certain message that

needs to be regulated and aligned with the public sphere. In the second part, I will argue that public works of art and the “artistic autonomy” argument also face limitations when their function is to honour or commemorate.

## I. INTRODUCTION

Controversies surrounding artworks are quite common; those surrounding public artworks are endless. The concept of public art is not new: church facades and cathedrals for instance are filled with sculptures and frescoes that are now crucial components of the history of art. Questions regarding the conditions for and the limits of public artworks have arisen as complex, non-representational or conceptual modernist public sculptures and monuments have increasingly seen the day, often leaving the public confused. As contemporary artists aim to be increasingly controversial in order to sell or to even be noticed, public artworks have become a battleground for controversies: how far can we take this modernist, avant-gardiste originality, specifically when it is on the streets? How can we even judge what is acceptable or not when it comes to artistic creation, a field supposed to be free of any rule or adherence to preexisting norms? Should we treat public artworks the same way we treat works in museums?

I will address these questions in this paper by arguing that there are certain limits to the immunity and artistic freedom that artworks benefit from. Caroline Levine, in her paper “*The Paradox of Public Art*”, makes the case for this complete artistic freedom specifically in the case of avant-garde public artworks, I will therefore explain and discuss what I shall call the “artistic autonomy” argument, in order to find its limitations. This argument, which will be clearly exposed after this introduction, grants immunity to artists because art should express their very own artistic or creative individuality: it is not for the people and thus should not be subjected to public opinion. The two following parts will argue that this argument has limits in the case of avant-garde works, due to the message conveyed by the artwork, and in the case of monuments and memorials.

## II. THE LIMITS OF ARTISTIC AUTONOMY: PUBLIC AVANT-GARDE

Art is pure creative and personal expression and for that reason, many agree that it should not be limited or silenced. Art has also always been full of controversies: showing what was considered too much skin on a royal portrait for instance surely brought the same uproar as Marcel Duchamp’s *Fontaine*. But both were eventually accepted, and even praised for their boldness. Thus, artists’ extravagances are difficult to criticize, and they are sometimes even preferred, especially in avant-garde art, which is now a term used for any artwork that shows radical originality and pushes the boundaries of contemporary artistic creation.

In her paper, Caroline Levine develops the idea of artistic autonomy or individuality as an argument for the preservation of avant-garde public artworks, which are rarely successful or appealing to the public, even though non-conventional art has been in circulation for decades.<sup>153</sup> This argument originated from a claim by the artist Richard Serra, whose public artwork *Tilted Arc* was removed because

---

<sup>153</sup> Levine, Caroline, “The Paradox of Public Art: Democratic Space, the Avant-Garde, and Richard Serra’s ‘Tilted Arc.’” *Philosophy & Geography* 5, no. 1 (2002): 54.

of general dislike. Serra's response to the controversies surrounding his work was to promote "the integrity of the autonomous artist", which implies that he made certain decisions regarding his artwork that only concern him, as an independent artist. According to him, artists should not try to please the public's mainstream taste, but they should only do works for themselves, to express their autonomous creativity, and this is what seems to grant them this immunity.<sup>154</sup> This argument raised by Serra and further supported by Levine, is what we will refer to in this paper as the "artistic autonomy" argument. So the question is: does the immunity granted by the "artistic autonomy" argument apply to public artworks?

This is an important question given that one of the fascinating things about art is that a work displays the artist's psyche through artistic practices. And often, the more tormented the artist is, or the more distant from normative beliefs, the more fascinating their work is. Furthermore, artworks are more than their aesthetic value, especially in public: they can send a message, or they can be required to fulfill a certain function for instance. Levine's artistic autonomy argument surely makes a pertinent claim that has always been evidently confirmed in the history of art. However, I wish to argue in this first part that applying this argument to public avant-garde artworks is unsuccessful, because public artworks have more than just an aesthetic value, compared to works of art displayed in private settings.

Although the artistic autonomy argument seems successful in the context of museums or art galleries, it becomes more than a question of aesthetic appreciation when the work is displayed in a public space and I will argue that, for this reason, this argument fails in the case of public avant-garde works specifically. Hilde Hein investigated the issue of what makes the meanings of public and private art different in his paper "*What is Public Art?*". Ultimately, he grants that public art is meant to show the "private vision of the individual artist"<sup>155</sup> while at the same time being, to a certain extent, a testimony of an institution's values and beliefs.

Since sharing the artist's private vision has never been considered an issue when their work is displayed in museums, why should we restrict public artworks? It is a given assumption that we do not question the acceptability of the artworks we see in museums because we cannot put restrictions on creativity, even when this creativity is too original for the public to enjoy or understand. This is because artworks in these private places solely have an aesthetic value: they are displayed for aesthetic appreciation only. Works in museums are here to see for the people who want it and arguably, they usually do not send any specific message from the institution for instance, because we know that museums only show these works for *voluntary* public appreciation. People in museums will be faced with works that promote all kinds of ideas, from immoral violent ones to revolutionary and anti-conformist ones. But since this is what a museum does, since these messages do not leave the walls of the gallery, it does not actually broadcast a message publicly. The message is simply here, for people to see, only if they desire. Public artworks differ in this manner: depending where this creativity is displayed, it can take up more than simply an aesthetic value, and it may need to be monitored in order to control the message that it sends to the viewer.

---

<sup>154</sup> Levine, "The Paradox of Public Art", 55.

<sup>155</sup> Hein, Hilde, "What Is Public Art." in *Arguing about Art: Contemporary Philosophical Debates*, (London: Routledge, 2007), 406.

When a work is displayed in public, although it does show the artist's psyche and creative individuality which is what makes it interesting, it also does more than that. If a work is commissioned to be displayed in public, it has been reviewed and imagined while keeping in mind that it is visible to everyone, that some people will walk by it on their way to work every single day. If it is often said that art has no specific purpose other than aesthetic appreciation, but arguably, public works do have a function. George Tsai develops this idea in "*The Morality of State Symbolic Power*" in which he argues that monuments, statues, and even street names constitute a form of expression by the state. He argues that these "stand for or represent something with political content favored by the state"<sup>156</sup>. In other words, infrastructures commissioned by a public institution (such as banks or corporations) or by the state, including works of art, constitute a symbolic state power in one way or another.

Although this has resulted in many successful works of public art, arguably this cannot be reconciled with avant-garde works in particular. By definition, avant-garde artists aim to go against the accepted societal norms and values of their time. They know the more they shock and outrage, the more successful they will be as avant-gardistes. The artistic autonomy of these artists that Levine praises in her paper can therefore be defined as a wish to go against what is commonly liked or known; in other words, to go against public approval and taste. Therefore, these types of public artworks convey a message. In the case of avant-garde works, the message conveyed by these works is their desire to go against public taste and be unconventional. I would argue that this is where the issue lies and where we need to consider the limits of public artworks: how can we expect the public to accept a work that aims to defy public taste? Of course one could say that many avant-garde works are enjoyed by the public, although usually only by the art-educated elite. But this is paradoxical: if an avant-gardist work is broadly enjoyed by the public, it becomes an unsuccessful avant-gardist work. In this context, if the message conveyed is problematic and paradoxical, this seems to tell us that there are *some* limits to this "artistic autonomy" because, as we have seen, it is not always fitting to the public, namely when it intends to go against public taste.

The Serra case, which Caroline Levine uses in her paper to support her argument, is a good example of this paradox. In 1981, Richard Serra created an avant-garde work called "*Tilted Arc*" to be placed in the Foley Federal Plaza in Manhattan. After eight years of debate and controversy, Serra's *Tilted Arc* was removed. The people who voted for its removal maintained that they did not like it for various reasons: it blocked the view, it created shade, and it was simply strange-looking. If Serra's goal all along was to go against public taste, how do we ever resolve such an issue? If nobody had contested it, then Serra would have failed as an avant-garde artist. But people disliked the artwork, and it is easy to understand why they did not want to walk by an artwork that not only aimed to dismiss their own "mainstream" taste and norms, but that they also found aesthetically unappealing. I will now consider a possible counter-argument to this claim.

Stating that art should try to be on par with the public's taste is of course controversial. A possible way to object to this statement is to point to the fact

---

<sup>156</sup> Tsai, George, "The Morality of State Symbolic Power" *Social Theory and Practice* 42, no. 2 (2016): 321.

that all the greatest artists were at one point described as shocking, too original or too complex for their contemporary audience. The first impressionists were considered scandalous, and the first cubists were considered mad. The most famous artworks nowadays were hated and criticized at first, but one could argue, and rightly so, that this is what makes art evolve and if we had put some limits on these artists' artistic autonomy, art would not be what it is today. I grant that this counter-argument is surely true, but I would still argue, as a way to respond to this potential objection, that it is different when applied to public artworks.

The reason for this is that when granting this counter-argument, it becomes problematic as it can justify displaying *anything* in public. If we think of the subject matters of famous artworks, many of them depict violence against women for instance, when others depict nude underaged girls. Many of Picasso's works were a way for him to paint away his rape fantasies, and Gauguin's iconic nude paintings of "Tahitian women" were really nude paintings of underaged girls, but this did not stop them from being admired today. Without commenting on the morality of artworks and artists, we can still use these examples to see that if we let all types of works of art to be exhibited in public, even shocking works, without questioning them because they will make the art world evolve or because the "greatest works were hated at first", it is crucial to consider where these works are displayed and what kind of message they convey. We have seen that works in private settings usually do not convey any type of message since they only have an aesthetic value, and that this is the reason why all the examples of artworks I mentioned are not removed from museums. However, when they are displayed outside, for everyone to see, these artworks become more than just aesthetic objects, as they convey a message publicly.

I will use Alberto Giacometti's sculpture entitled *Woman with Her Throat Cut* as an example here. This bronze sculpture represents a woman, who looks very similar to an insect, lying on the ground with, as the title indicates, her throat cut. This sculpture is explicitly a form of representational violence towards women, but this is not questioned as the work is currently in a museum, a private space which displays artworks for aesthetic enjoyment only, with no intention of spreading some type of message about violence towards women. However, as we have previously seen, it would constitute more than a simple aesthetic object if it were to be displayed in public. Like any public sculpture, it would carry some kind of message, like Tsai has argued, and this becomes problematic considering the artwork's subject matter: a woman with her throat cut. This constitutes a response to the potential counter-argument raised earlier according to which all greatest works of art were hated or found shocking at first and this is why we should not put limits on artistic autonomy. Indeed, as we have seen, that this cannot apply for every case and especially not in the case of public art. As a result, the subject matter of a public work of art needs to be monitored, which confirms that there are some limits to artistic autonomy.

Levine mentions in her paper that the goal of avant-garde artists is that one day, the public sphere will be like a museum, and artists will be able to expose their works as they please, and have full creative autonomy just like works in museums do.<sup>157</sup> I have argued in this first part of my paper that this is not realistic, and that there are certain limits to this artistic autonomy defended by Levine and

---

<sup>157</sup> Levine, "The Paradox of Public Art", 2.



Serra. I have shown that this is specifically the case with regards to public works of avant-garde, since the message these works convey is not fitted for the public. This is not to say that all works of public art should be accessible or enjoyable by the public, because it would be impossible and dull to make a work that appeals to everyone. It simply shows that these are cases in which artistic autonomy cannot be without limits or restrictions. I will argue in the following part that these restrictions also apply to other types of public artworks: monuments and memorials.

### III. THE LIMITS OF ARTISTIC AUTONOMY: MONUMENTS AND MEMORIALS

We have established that public avant-garde artworks convey a message publicly that has to be monitored and, to a certain extent, aligned with public beliefs and norms. I will argue in this second part that these limitations in terms of artistic freedom or autonomy also apply to public artworks meant to honour or commemorate something or someone. These types of public works are perhaps the most meaningful for a population or a nation, as they reinforce shared values and commitments and they act as a reminder of history. However, a consensus is almost never attained on whether they share the intended message the right way. I will therefore show that monuments and memorials constitute another type of limitations that public art needs to navigate, because there is more to these works than their aesthetic value and the message shared by these works is not for the artist but for the people or for the state in general. Thus, this message needs to be understandable and accessible to the people, not only to the art-educated elite.

Hilne Hein maintains in his paper that the modernist aesthetic is centered on the individual rather than the public. Modernist and avant-garde artists share their own visions or ideas, and make them difficult to understand. Nowadays, works of art in galleries are not as accessible as they used to be, as they invite some kind of retrospection and reflection. This is precisely what these artists want, to make people work for meaning and for appreciation. So how can modernist artists -even the ones who are not necessarily part of the avant-garde but who simply don't make figurative works- reconcile their desire for complexity with the attempt to make a successful memorial, understandable and enjoyable by an entire population? This is where it seems like public artworks face limitations once again.

In his paper "*How Memorials Speak to Us*", Geoffrey Scarre explores the meanings of commemorations and how the public interprets these meanings. This meaning-making activity is crucial when it comes to public artworks meant to commemorate, as they have a function to fulfill: honouring something or someone and making sure that the public understands it. According to him, monuments have an "illocutionary" function, which constitutes the message that is intended to be shared by a memorial. But they also have a "perlocutionary" effect which is the actual effect that the monument has on the viewer. This effect can be intended or not, depending on how the viewer interprets the illocutionary function of the memorial.<sup>158</sup> Scarre argues that the perlocutionary effect of these works can and should be monitored, and that the success of a memorial can be

---

<sup>158</sup> Scarre, Geoffrey, "How Memorials Speak to Us" *Philosophical Perspectives on Ruins, Monuments, and Memorials*, (2019): 21.

assessed depending on how successful it is in its illocutionary function. This means two things: that monuments convey meanings by being non-verbal speech acts, and that this meaning is not guaranteed to be understood correctly by the viewer, which could result in an unsuccessful act of honouring from the monument.

This argument helps to support my claim that when a work of public art is supposed to honour something or someone, there are limits to the artist's autonomy as its illocutionary function should try to be aligned, as much as possible, with the perlocutionary effect it will have on the viewer. In other words, these types of public artworks should not be too hard or complex to understand, they should not necessitate any type of art education. The reason for this is that, unlike "regular" artworks which often have open meanings and invite the viewer to reflect on their own interpretation and enjoyment of the work, memorials have two functions that go beyond aesthetic enjoyment: they must be for the people and they must be representative of the institution or the nation's values and ideals. Ultimately then, for an artwork to succeed in its illocutionary act of honouring, viewers have to recognize the conventions of honouring in which it is situated. If it is so avant-garde that those conventions are not recognized, it will not be a successful speech-act of honouring. If, according to Scarre, a successful memorial depends on how well-interpreted its illocutionary function is, it will have to be clear to all what the work is commemorating, which constitutes a restriction on artistic autonomy.

As an example, we will look at the case of the *Lin Vietnam War Memorial*, which was one of many controversial public works of art. The message that the artist's work was supposed to convey (its illocutionary function) -which was to honour the Vietnam War veterans- did not perfectly align with how the public interpreted it (its perlocutionary effect). In 1982, the 21-year old college student Maya Lin won the national competition for the design of the Vietnam Veterans Memorial in Washington, DC, for which the judges were veterans and art experts. Although she won the competition fairly, many veterans were quick to protest the sculpture. The reasons for the dislike of the work were quite varied: some thought it did not highlight the heroism of the war, some did not like the fact that it was black, some found it depressing that the work sinks into the ground as one walks by it instead of standing tall. Overall, the people who expressed their discontent with the work did not feel honoured by it, calling it "the black gash of shame" and criticizing its lack of narrative, which made it look like the names on the monument could have been victims of a bus accident instead of war heroes. Here, we can therefore see the discrepancy between the illocutionary function of the monument and the perlocutionary effect it actually had. Although the artwork still stands today and people eventually came to terms with it, its minimalism made the monument's illocutionary function too difficult to identify, resulting in what people deemed to be an unsuccessful act of honouring. The controversy around this memorial supports the claim that when it comes to commemorating, the artist cannot have complete freedom and autonomy if it wants the monument to be successful in its act of honouring; Unlike any other work of public art, memorials are for the people, not for the artist, which means that it needs to be accessible and understandable to the viewers before being a means of expression for the artist. I have thus shown that memorials and monuments also constitute an example of limitations to the artistic autonomy argument.

Since there are so many controversies and debates around this topic, there are also many counter-arguments to this claim. Thus, a possible opposition could find that art that is too easy to understand and that pleases everyone is “bad” art. This is a claim that is quite popular in aesthetics; artworks that are too figurative or too easy to understand lack the enjoyment that comes from the reflection when faced with an artwork. It can even become “kitsch”, a term largely used to describe artworks with no depth or no variety of interpretation, such as a cute kitten playing or a baby crying. These works are figurative, easy to understand and realistic: their illocutionary function and the message they are trying to convey to the viewer are explicit, but ultimately, they are considered to be bad art. So according to this counter-argument, a successful monument -one with a clear, understandable message- can only be a bad work of art. So do we need to settle for bad art in order to make a monument successful?

I would respond to this counter-argument that this is not necessarily the case. First of all, if most art considered “kitsch” is usually unreflective, not every unreflective artwork is kitsch. Simply because art has evolved in a way that promotes meaning-making and complexity through very little visual evidence does not mean that any work that depicts explicitly what it wishes to convey to the viewer is bad or kitsch. For instance, a memorial representing a fighting soldier with the names of war veterans -such as the Marine Corps War Memorial or the Korean War veterans memorial- makes the subject matter and the message shared by the sculpture very explicit, while not necessarily being kitsch or bad. Just because a work of art’s subject matter is harder to understand or interpret does not mean that less reflection will come out of one’s experience with it. We can be faced with a very realistic and figurative sculpture for instance and still find deeper meaning, the same way that we can face an avant-garde sculpture composed of nothing but blocks and see absolutely nothing. The issue is, therefore, not whether the work is reflective or not, but whether it will be understood correctly as a commemoration. This requirement for explicitness thus appears to be crucial for memorials, as they have a specific function and engagement towards the people and the public institutions that it represents. So, this constitutes another limitation of Levine and Serra’s artistic autonomy argument.

To further support this argument and highlight the importance of making the illocutionary function of public memorials explicit, I will mention a last example which is currently discussed extensively in the news: the Mary Wollstonecraft Memorial in London. The sculpture, made by the avant-garde artist Maggi Hambling, has been receiving unrelenting criticism since it was recently inaugurated. The memorial represents a naked woman, who is clearly not Mary Wollstonecraft, standing on an abstract-looking silver mount; the plaque reads “for Mary Wollstonecraft”. It is unclear why the artist chose not to depict the woman for which the monument was commissioned, but this constitutes an adequate representation of this second part’s argument. Treating it the same way as all her other “regular” non-commemorative or public works of art, the artist has decided to follow her artistic autonomy: in an interview for *The Guardian*, the artist replied to criticism by saying that with her art, she needs “complete freedom.”<sup>159</sup>

---

<sup>159</sup> Vanessa Thorpe, “‘I need complete freedom’: Maggi Hambling responds to statue critics,” *The Guardian*, November 14, 2020, <https://www.theguardian.com/artanddesign/2020/nov/14/i-need-complete-freedom-maggi-hambling-responds-to-statue-critics>

Although this is what art is about, this controversy further proves my point that this complete artistic freedom does not apply to all works of art, specifically with public artworks. Here, the artist has decided not to figuratively and explicitly honour Mary Wollstonecraft, thus giving us an example of a non-figurative, yet bad work of art, according to her critics. This has resulted in general confusion and dissatisfaction from the public, who claimed that this sculpture simply does not honour the memory of the 18th century philosopher, but constitutes a way for the artist to express herself. However, this was not the intended function of this specific work: the nature of the artwork, even if it is aesthetically interesting, seems inconsistent with the conventions of honouring and thus does not succeed to honour Mary Wollstonecraft. So, this example proves the importance of making the illocutionary function of public memorials explicit, even if it means putting limits on artistic freedom and autonomy.

#### IV. CONCLUSION

To conclude, I have argued that some restrictions can, and sometimes should, be put on artistic autonomy. While I agree that art is fascinating in the fact that it gives complete freedom to the artist, allows the public to see through their mind, and interpret it however they like, there are restrictions to such autonomy in the case of public art. I argued in my first part that this autonomy finds its limits in public avant-garde and, in my second part, that it also does so in the case of memorials or monuments.

It is worth mentioning that it is unrealistic to demand that art sparks no controversies or disagreements. Part of what makes art interesting is that elicits constant controversies and will likely continue to do so in the future. This has become especially true now that we value artworks not only for their aesthetic qualities but also for their individualism and their ability to challenge pre-existing artistic norms. And, although individualism is admirable and necessary as an artist, it becomes a much more complex issue when the public is involved, and this is something that artists need to take into consideration.

BIBLIOGRAPHY

Kelly, Michael. "Public Art Controversy: The Serra and Lin Cases." *The Journal of Aesthetics and Art Criticism* 54, no. 1 (1996): 15–22. <https://doi.org/10.2307/431677>.

Levine, Caroline. "The Paradox of Public Art: Democratic Space, the Avant-Garde, and Richard Serra's 'Tilted Arc.'" *Philosophy & Geography* 5, no. 1 (2002): 51–68. <https://doi.org/10.1080/10903770120116831>.

Neill, Alex, Aaron Ridley, and Hilde Hein. "What Is Public Art." Essay. In *Arguing about Art: Contemporary Philosophical Debates*. London: Routledge, (2007); 406-13.

Scarre, Geoffrey. "How Memorials Speak to Us." *Philosophical Perspectives on Ruins, Monuments, and Memorials*, (2019): 21–33. <https://doi.org/10.4324/9781315146133-3>.

Thorpe, Vanessa. "'I Need Complete Freedom': Maggi Hambling Responds to Statue Critics." *The Guardian*, November 14, 2020. <https://www.theguardian.com/artanddesign/2020/nov/14/i-need-complete-freedom-maggi-hambling-responds-to-statue-critics>

Tsai, George. "The Morality of State Symbolic Power." *Social Theory and Practice* 42, no. 2 (2016): 318–42. <https://doi.org/10.5840/soctheorpract201642218>.

