

Logos : λόγος

Volume XVIII • Spring 2022

Copyright © 2022
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 XVIII • Spring 2022

Editor-In-Chief:

Leonardo Villa-Forte

Club President:

Liam Galey

Vice President:

Jacqueline Allen

Treasurer:

Meghana Gavirneni

Secretary:

Sam Vučić

General Editorial Staff:

Alkis Boukas

Fabio Cabrera

Adam Chih

Shriya Desai

Sophia Gottfried

Ben Knepper

Rachel Kodysh

Ethan Kovnat

Ben Dever-Mendenhall

Alejandro Maris Natera

Xiaoyi Wu

Zelai Xu

Tracy Lihui Zhang

Faculty Advisor:

Harold Hodes

Contents

Editors' Introduction	6
Forgiveness: A Descriptive Analysis <i>Kenneth Quek</i> <i>University of Warwick</i>	7
A Phenomenological Experience of the Novel <i>Virginia Moscetti</i> <i>Swarthmore College</i>	25
How Wide is the Mind? The Extended Mind, the Narrow Mind, and Critical Disability Theory <i>Kexuan Liu</i> <i>University of Virginia</i>	37
Split-Brains, Mental Unity, and the Combination Problem for Panpsychism <i>Robison Yost</i> <i>Wright State University</i>	45
Plato and the Problem of Tragedy <i>Weiming Sheng</i> <i>Vassar College</i>	59

Editors' Introduction

This year represented a gradual and yet enthusiastic return to normalcy for our vibrant philosophical community, marked by fruitful philosophical discussion among our editorial board in search of the papers we present to you in this edition. It is with great pleasure that we, the Logos editorial board, present the eighteenth edition of *Logos: The Cornell Undergraduate Journal of Philosophy*.

We received 91 submissions from institutions across the world. From these, half advanced to our second round of review and 16 advanced to our final round. After careful consideration, we finally selected the top 5 papers we deemed deserved publication. We determined that each of these impressive papers were highly competent and well-written pieces of undergraduate philosophical thought. In particular, they demonstrated understanding of the field they contributed to and great clarity and precision of reasoning, all while discussing interesting and important topics from original perspectives. Each of them represents some of the highest quality of work that serves as evidence for a thriving undergraduate philosophical community. Therefore, we are grateful for the opportunity to present these works for your reading pleasure.

The topics you'll encounter below include forgiveness, phenomenology, philosophy of mind, and finally ancient philosophy. We'd like to thank the authors whose writing we are showcasing. In particular, we are proud to bring to our readers such a wide range of philosophical topics. As in the past, each submission received personalized feedback from one of our editors, including the papers that were not selected for publication. We hope all of the authors will continue to develop and share their philosophical ideas with us and the world in the future.

In particular, I would like to thank the entire editorial board that worked tirelessly to produce another edition of *Logos*. This year's edition would not be possible without their commitment to philosophical discussion, openness to all viewpoints, and professional dedication to immerse themselves in all 91 philosophical arguments. I'm especially grateful to the Sage School of Philosophy for their funding and support, in particular to Pamela Hanna and Dorothy Vanderbilt for helping our editorial staff in any way necessary. Finally, thank you to our faculty advisor Professor Harold Hodes for supporting our journal.

Leonardo Villa-Forte
Editor-in-Chief

Forgiveness: A Descriptive Analysis

*Kenneth Quek
University of Warwick*

FOREWORD

Editors Sam Vučić, Ben Dever-Mendenhall

Describing the nature of forgiveness has proven a surprisingly complex philosophical endeavor, but this paper manages the task with elegance and concision, inviting readers to question their own assumptions about what it means both to forgive and to be forgiven. We appreciate the clarity with which Quek presents his arguments and the earnesty with which he grapples with questions so fundamental to the human experience. “Forgiveness: A Descriptive Analysis” features all of the hallmarks of a robust undergraduate philosophy paper – a clear thesis, a logical structure, and a series of arguments accessible to readers less familiar with background literature. We also enjoyed the paper’s nuanced counterarguments, which extend Quek’s analysis beyond the scope of superficial discourse.

“Forgiveness: A Descriptive Analysis” contains a rich, interdisciplinary reconciliation of competing accounts which seek to understand *true forgiveness*: that nebulous “thing” which so obviously holds *unspeakable import* in our ability to harmoniously bond with another individual. Quek’s work was met with acclaim by our editorial board because of its comprehensiveness, sophistication in its communication, and delicate acuity in presenting the relational account. In virtue of these and other qualities, “Forgiveness: A Descriptive Analysis” claimed the top prize of this year’s edition of our publication.

Dear reader, whether you are a clinical psychologist, devoted partner, or honest friend, know that essential lessons lie in wait. Quek goes beyond semantics, beyond abstraction, offering instruction in how to preserve, come unescapable wrongdoing, perhaps our most invaluable treasure: *the love of and for our fellow beings*.

ABSTRACT

Forgiveness makes up a significant portion of our everyday moral lives. Yet, it is a nebulous concept, ambiguous and confusing at best and paradoxical at worst. With contemporary philosophy’s growing interest in the subject, multiple accounts of forgiveness have surfaced to attempt to resolve these tensions and create a descriptive analysis of forgiveness. This paper first examines the current paradigmatic view to the question, dubbed the ‘Standard View’ – namely, that forgiveness is a kind of reactive attitude. The discussion reveals that the relational aspect of forgiveness is central to any descriptive analysis of forgiveness. The paper then uses this conclusion to advance a novel account of forgiveness, labeled ‘The Relational Account’, that conceives of forgiveness as a process of repairing relationships damaged by wrongdoing. Under this view, forgiveness cannot be achieved unilaterally, through one single person’s actions, but rather emerges through a cooperative process of reconciliation.

I. INTRODUCTION

Most traditional Western theories of normative ethics focus on the conditions under which an action may be considered morally justified, but tell us little about how to act after wrongdoing or being wronged. This means that forgiveness, which deals with ‘post-wronging’, does not find an easy place in our traditional moral

theories. Yet, given how significant forgiveness is in our everyday moral lives, an account of forgiveness must be found for any holistic account of morality. With the increasing level of interest in the topic of forgiveness in contemporary philosophy, the need for a robust description of forgiveness is now more relevant than ever. This paper aims to establish a groundwork for such investigation by describing what forgiveness is, and what it accomplishes. It evaluates the current paradigmatic view to the question, dubbed the ‘Standard View’ – namely, that forgiveness is a kind of reactive attitude. Upon establishing that this view is lacking, the essay constructs an alternative account – that forgiveness is a repairing of relationships damaged by wrongdoing. This paper then proceeds to investigate how this repairing is achieved.

It is first important to establish in more precise terms what the paper seeks to do. Some philosophers have created descriptions of forgiveness wherein they describe what it is for someone to be forgiving – in other words, they sketch a picture of the virtue of forgiveness¹. Such discussions are not this paper’s interest. Unlike descriptions of the virtue of being forgiving, this paper concerns itself with the *act* of forgiving; it explores what it is to forgive someone. More specifically, this paper focuses on what, fundamentally, forgiveness is - it aims to describe what obtains uniquely for forgiveness such that a description of forgiveness sketches some distinctive phenomenon. As such, one may formulate the question this paper seeks to answer in the following manner:

Essentially, what is involved in forgiving?

The paper will progress as follows. First, it will sketch a broad outline of the view that, fundamentally, forgiveness is a reactive attitude or change in reactive attitude. This will be referred to as the Standard View. The paper proposes a challenge to this view. Then, the paper explores a more specific variation that aims to dissolve the challenge to the Standard View. Some limits to this variation will be explored. Then, the paper proposes that these limits can be bypassed by adopting the view that forgiveness consists of some relational phenomenon – that is, that it is some way in which one relates to other people. This shall be referred to as the Relational Account. By analysing a model of forgiveness aimed at understanding grave wrongs, this paper extracts key insights that it generalises into the Relational Account in order to describe what exactly the relationship forgiveness consists of might be.

II. THE STANDARD VIEW

Contemporary philosophical literature on the nature of forgiveness is sparse, as interest in the subject is relatively new. However, there has still emerged a family of views that one may take to be mainstream, or the most popular. According to these views, to forgive someone is to adopt some attitude or a change in attitude. This family of views will be referred to broadly as the Standard View. Despite this paper treating such views as one single family, it is important not to consider them monolithic – many philosophers who can be said to hold some form of the Standard View disagree with each other on important specificities of what

¹ Thomason, “Forgiveness or Fairness,” 233; Allais, “Elective Forgiveness”, 637.

forgiveness is. By way of illustration, there exists substantial debate on what the attitude referred to in the Standard View consists of. For Griswold, forgiveness is a reactive attitude that involves overcoming another reactive attitude, which is that of resentment.² However, some philosophers argue that forgiveness admits of a plurality of emotions to which forgiveness reacts to, including hatred, indifference, sadness, etc.³ Even within the view that forgiveness is a reaction to resentment, there is disagreement on what precisely resentment entails. Among a plethora of other perspectives, some believe that it is a 'a kind of protest', while some others view it as 'a feeling of insult.'⁴ Beyond the nature of resentment, still more variance can be found in how advocates for the Standard View see the relationship between the act of forgiveness and its substrate emotion (resentment or otherwise). Some view it as an 'overcoming' – that is, when the relevant emotion is 'replaced on a spontaneous level with goodwill toward the offender.'⁵ Still others treat it as merely an 'elimination' of the relevant emotion, the crucial difference in the two perspectives being that the latter admits of cases where the relevant emotion is not replaced with goodwill.⁶ Given the broadness with which the Standard View in this paper is being defined, the set of views which may be considered as belonging to the Standard View is diverse at most, if not all, levels at which they discuss the nature of forgiveness.

However, despite the diversity of these views, they are all in agreement that forgiveness is primarily and fundamentally a mental affair – it is some way of thinking, or some process that occurs in the mind. For clarity, it is this specific proposition that this paper refers to as the Standard View.

The adoption of this view seems to create contradictions in our account of forgiveness when considered in relation to the original wrong that one seeks to forgive. This problem is often referred to as the paradox of forgiveness, or Kolnai's paradox. It is as follows:

'... [Either] forgiveness is objectionable and ungentle inasmuch as there is no reason to forgive, the offender having undergone no *metánoia* ('Change of Heart'), but persisting in his plain identity qua offender ... [or] at the other end of its spectrum, forgiveness seems to collapse into mere redundancy, or the mere registering of moral value in the place of moral disvalue.'⁷

To paraphrase, we first begin with the underlying assumptions of the paradox. For some person S1 to forgive another person S2, S2 must have committed some wrongdoing. This is uncontroversial, and this paper will not seek to challenge this assumption. At this juncture, Kolnai presents readers with a dilemma: either forgiveness is a condonation of the wrongdoing, or it is not. If forgiveness is a condonation of the wrongdoing, then forgiveness must always be unjustified, because it is wrong to condone wrongful actions. If forgiveness is

² Charles Griswold, *Forgiveness: A Philosophical Exploration* (Cambridge: Cambridge University Press, 2007), 39.

³ Jeffrie Murphy, *Getting Even: Forgiveness and Its Limits* (Oxford: Oxford University Press, 2003), 59.

⁴ Pamela Hieronymi, "Articulating an Uncompromising Forgiveness," *Philosophy and Phenomenological Research* 62, no. 3 (2001): 530; Jeffrey Blustein, *Forgiveness and Remembrance* (Oxford: Oxford University Press, 2014), 33.

⁵ Margaret Holmgren, *Forgiveness and Retribution* (Cambridge: Cambridge University Press, 2012), 61.

⁶ Paul Lauritzen, "Forgiveness: Moral Prerogative or Religious Duty?," *The Journal of Religious Ethics* 15, no. 2 (1987): 143.

not a condonation of the wrongdoing, then, for any non-circular reason for why forgiveness is not a condonation of the wrongdoing, that same reason may be used to explain how S1's 'change of heart' towards S2 is morally justified, without appealing to the concept of forgiveness at all. Therefore, either forgiveness is always morally unjustified, or conceptually redundant.

This paradox is particularly damaging for variations of the Standard View. This is because the Standard View, by virtue of treating forgiveness as solely some mental activity, has no easy response to the paradox. Views that do not appeal to forgiveness as some mental activity, by contrast, may sidestep Kolnai's paradox by asserting that forgiveness does not even describe 'changes of heart' to begin with – they map onto another kind of phenomenon. Put more clearly, such views would assert that i) forgiveness is not a condonation of wrongdoing, but also that ii) what forgiveness describes is distinct from a justification of S1's 'change of heart' towards S2, as 'changes of heart' refer to attitudes while forgiveness does not. Such a strategy is not available to proponents of the Standard View, as they are committed to the view that forgiveness is mental in nature (and, more often than not, that forgiveness does in fact describe changes of heart). Kolnai's paradox thus remains a significant obstacle that all versions of the Standard View must account for.

III. THE HAMPTONIAN ACCOUNT

However, the Standard View is not defenceless against the paradox. In *Forgiveness, Resentment, and Hatred*, Jean Hampton makes a case for a version of the Standard View that bypasses such challenges.⁸

The strategy Hampton employs is to draw a distinction between reactive attitudes towards actions and reactive attitudes towards people. She does so by conducting a semantic analysis of four different reactive attitudes, which this essay reconstructs into the following matrix:

	Personal	Moral
Towards actions	Resentment	Indignation
Towards people	Simple Hatred	Moral Hatred

All four of these attitudes are a kind of anger. However, they differ in what the anger is directed towards, and where this anger stems from. Indignation, for example, stems from a sense of aversion towards some immoral cause or practice, and is directed towards the immoral action itself. On the opposite end of the matrix, Simple Hatred stems from a protest against demeaning treatment towards oneself, and is directed towards the person who demeaned them.⁹

She further elaborates on his account by establishing links between the reactive attitudes. Resentment may lead to a simple hatred towards the thing which does the actions one resents; for example, my resentment for rudeness

⁷ Aurel Kolnai, "Forgiveness". *Proceedings of the Aristotelian Society* 91, no. 1 (1973): 97.

⁸ Jean Hampton, *Forgiveness and Mercy* (Cambridge: Cambridge University Press, 1988).

⁹ *Ibid.*, 58-61

may lead me to develop a simple hatred for rude people. Likewise, my indignation towards a moral wrong may lead me to develop a moral hatred for someone; if I knew someone who was an abusive spouse, I may come to hate him for his moral deficiency¹⁰.

Given this distinction, Hampton's resolution to Kolnai's paradox becomes fairly straightforward. In line with the Standard View, forgiveness pertains to attitudes, but it specifically pertains to reactive attitudes towards people. Under Hampton's view, the act of forgiveness is characterised by an overcoming of moral hatred, but does not entail forswearing resentment or indignation towards the wrongdoer's actions.¹¹

When S1 forgives S2, S1 is not condoning S2's wrongful behaviour, because S1 still possesses resentment or indignation towards the action. However, characterising such an action as forgiveness is still not redundant, because forgiveness refers specifically to the change in reactive attitude S1 has towards the S2 themselves, a phenomenon that can still paradigmatically and usefully be called forgiveness. Forgiveness does not entail that one revise one's indignation towards that which wronged you, or even indignation at someone's bad character traits. It is a revision of one's evaluation of the person themselves – where this person is understood as something other than or more than the character traits which one does not approve. This distinction between actions and persons-in-themselves is how Hampton avoids Kolnai's paradox.

IV. RESPONSE TO THE HAMPTONIAN ACCOUNT

This paper accepts that Hampton adequately resolves Kolnai's paradox. However, it also argues that Hampton's view, as it stands, is still too problematic to be considered an adequate account of forgiveness. This section will proceed as follows: first, it presents an argument for why Hampton's view is not necessary for an account of the act of forgiveness. It will then demonstrate that it is also not sufficient for an account of the act of forgiveness.

First, it would be useful to state in explicit terms what Hampton's view would assert is essential for some act to count as an act of forgiveness. Call this the Hamptonian Criterion.

Hamptonian Criterion:

What is essentially involved in S1 forgiving S2 is S1 giving up their moral hatred towards S2 (not necessarily giving up their sense of indignation or resentment at S2's actions).¹²

According to the Hamptonian Criterion, forgiveness necessarily involves that S1 revises their evaluation of S2's character, thereby undergoing a 'change of heart' and giving up their moral hatred for S2.

¹⁰ Despite making a distinction between personal and moral protest, Hampton does not think they are mutually exclusive – one may feel indignant at some action but also feel personally offended at the action. The distinction lies specifically in what it is that these reactive attitudes are protesting, rather than what they strictly constitute. See Hampton, "Forgiveness and Mercy," 61 for more information.

¹¹ *Ibid.*, 83

¹² It is unclear whether Hampton intends for her account to be both necessary and sufficient for some act to count as an act of forgiveness. However, this paper assumes it to be so, for the purposes of demonstrating that it is neither.

However, it is dubious whether a 'change of heart' is even necessary for some act to count as an act of forgiveness in the first place. It seems that, in cases where S2 is already in good standing with S1, S1 may forgive S2 without ever revising their evaluation of S2. In order to demonstrate this, this paper compares two cases of forgiveness, each progressing in four phases.

Scenario 1:

- i) S2 wrongs S1
- ii) S1 develops moral hatred for S2
- iii) S1 learns something about S2
- iv) S1 forgives S2

Scenario 2:

- I. S1 learns something about S2
- II. S1 develops a good impression of S2
- III. S2 wrongs S1
- IV. S1 forgives S2

Oftentimes, one forgives someone not because they have performed any sort of special atonement, but because one comes to realise and empathise with the motivations or reasons behind that person's wrongdoing. As an example, one may come to forgive a friend's offensive behaviour if they come to realise that this friend is going through some emotionally turbulent experience. This does not mean they condone the offensive behaviour, but that they understand that said behaviour is not a holistic and accurate representation of their friend's character. This is the type of case that Scenario 1 aims to capture.

Scenario 2 reimagines the case, but now with one first learning that their friend is going through some emotionally turbulent experience before their friend had behaved offensively. If one forgives their friend in Scenario 2 cases, it seems quite implausible that this would be any less a case of forgiveness than that in Scenario 1 cases. It would be odd to think the timeframe that one comes to understand the context behind wrongdoing would have any effect on whether one's forgiving the wrongdoer actually counts as forgiveness. In Scenario 2 cases, it seems that one may forgive someone *despite*, or even *because*, their positive evaluation of them had not changed.

Thus, there exists some cases of forgiveness (such as Scenario 2 cases) which the Hamptonian Criterion, as it stands, would wrongly determine are not cases of forgiveness. Therefore, the Hamptonian Criterion as it stands is not a necessary condition for forgiveness.

Having established that the Hamptonian Criterion is unnecessary for an account of forgiveness, this paper now proposes another challenge to the view – that it is not sufficient for characterising forgiveness, by virtue of its inability to distinguish between who is capable of forgiving.

There are cases where one may maintain one’s evaluation of someone as ‘still decent’, or revise one’s evaluation of someone as ‘still decent’, in spite of that person’s wrongdoing, without this act counting as an act of forgiveness. Consider the following scenario.

Some way or another, I have come to know about some bully in a French middle school who punched their classmate. Believing that punching classmates is wrong, I disapprove of their actions and am indignant at the action. Being indignant, I develop a feeling of moral hatred towards the bully. However, I later learn that they come from a troubled family background, and struggle with socialising in a school setting due to their traumatising home life. Bearing this information in mind, I empathise with the bully, revising my evaluation of their character such that I no longer bear moral hatred towards them.

Here, it seems that the Hamptonian Criterion is satisfied – there was some wrongdoing that I was made aware of, and – after learning more about the bully - I gave up my moral hatred for them, despite still being indignant at their bullying. Yet, it would seem that this is not a case of forgiveness at all, because I do not have a personal relation to this bully. As some writers on forgiveness would note, I do not have the appropriate ‘standing’ with the bully to forgive them.¹³ I cannot forgive them, because there is nothing on my part to forgive, even if I have some feeling of indignation towards their actions. Thus, despite my having fulfilled the conditions of the Hamptonian Account, I have not forgiven the bully, because I am not able to in the first place.¹⁴

V. TOWARDS A RELATIONAL ACCOUNT OF FORGIVENESS

The preceding challenges posed to the Hamptonian Account cast doubt on whether Standard View theories are adequate descriptions of what, essentially, is involved in forgiving. For a satisfactory descriptive analysis of forgiveness to be obtained, a new account of forgiveness must be conceived – one that, this paper argues, conceives of forgiveness as essentially about relationships, and their repair.

In order to begin creating this Relational Account of Forgiveness, this essay condenses the insights gleaned from the analysis of the Hamptonian Account. The goal of this exercise is to establish guidelines for what conditions the new account must satisfy in order to avoid the limitations of its predecessors.

¹³ Brandon Warmke, *Forgiveness and its Moral Dimensions* (Oxford: Oxford University Press, 2021), 6.

¹⁴ It should be noted that Hampton did acknowledge standing as a necessary condition for forgiveness (Hampton, “Forgiveness and Mercy,” 21). However, it is unclear how her conception of forgiveness, which treats it as merely a forswearing of moral hatred, can explain why standing is necessary in the first place. This paper asserts that the necessity of standing suggests that forgiveness essentially involves relationships in a way not adequately captured by Standard View theories.

The Hamptonian Account falters in its ability to differentiate between people who are in appropriate standing to forgive and those who do not – relevantly, it struggles in differentiating cases of genuine forgiveness from those where one cannot forgive the wrongdoer despite evaluating their character as still decent.

This observation raises interesting questions. Returning to the bully example, it is clear that I did in fact revise my evaluation of the bully's character. It is equally clear, then, that my initial moral hatred for the bully had begotten something that is more positive in nature. Yet, as the paper has established, this phenomenon cannot be characterised as forgiveness. What, then, is this phenomenon?

Because of the existence of this phenomenon, we may conclude that moral hatred, when given up, may result in at least two things – forgiveness, and the phenomenon in question. We shall refer to this phenomenon as 'pardoning'. Upon understanding the context behind the bully's actions, and coming to empathise with the bully, one comes to a point where they are capable of pardoning the bully. Similarly to forgiveness, this is distinct from condoning what the bully has done, and thus also avoids Kolnai's paradox. Reusing Hampton's distinction, one may pardon the bully, recognising that they may still be decent despite having done wrong, but refrain from condoning the bully's actions themselves.

Thus, pardoning and forgiving seem to be similar in some regards. However, it would be difficult to deny that forgiveness is conceptually distinct from pardoning. The first way that they differ is that forgiveness requires that the subject is in an appropriate standing to forgive, whereas no standing is required for someone to pardon a wrongdoer. The second, and perhaps more interesting way, is that forgiveness fulfills a function within relationships that pardoning does not (as pardoning does not necessarily take place against the context of a relationship to begin with). It is this difference that motivates us to believe that forgiveness is essentially a phenomenon that occurs within relationships.

VI. THE EMERGENT MODEL OF FORGIVENESS

In order to begin constructing such a conception of forgiveness, this paper draws insights from Carse and Tirell's *Forgiving Grave Wrongs*. They present the Emergent Model of forgiving, in response to what they call the Classical Model's shortcomings in describing forgiveness in the case of world-shattering wrongs. The principles behind the Emergent Model will then be generalised and used in the formation of the Relational Account of Forgiveness in the following section.

It is not the focus of this paper to discuss the distinction between the Classical Model and the Emergent Model. As such, discussion and mention of the arguments presented by Carse and Tirell against the Classical Model will be kept to a minimum. However, a brief description of the Classical Model shall be provided as context for the Emergent Model.

The Classical Model is described as follows:

Under the Classical Model, 'the victim extends forgiveness to the perpetrator as a result of a deliberate decision following a morally reparative transaction – a transaction aimed at forgiveness.'¹⁵

¹⁵ Alisa L. Carse & Lynne Tirell, "Forgiving Grave Wrongs", in *Forgiveness in Perspective*, ed. Christopher Allers & Marieke Smit (Amsterdam: Rodopi Press, 2010), 45.

The paradigmatic example of a Classical Model theory of forgiveness, as Carse and Tirell note, would be certain Standard View theories, such as Griswold's theory, under which an offender's request for forgiveness, actions constituting that request, and the victim's decision to forgive based on a variety of criteria are emphasised.

At this juncture, Carse and Tirell propose a challenge to Classical Models in the form of world-shattering wrongs. These kinds of wrongs, such as genocides, are kinds of wrongdoing performed on a scale that destroys a victim's sense of reality and meaning.¹⁶ In such cases, a victim's 'moral orientation' is shattered to the point where they cannot even make sense of a perpetrator's offerings and requests for forgiveness to begin with – as Wiesenthal laments, 'it is impossible to believe anything in a world that has ceased to regard man as man.'¹⁷

This collapse of moral orientation destroys any normative framework that a victim and a perpetrator may share. World shattering wrongs rob victims of recognition of their basic moral status, and thus undermine the possibility of any 'transaction' of forgiveness. If the Classical Model were to be believed, forgiveness after world-shattering wrongs would then be impossible.

In order to make sense of forgiveness after world-shattering wrongs, Carse and Tirell reject the Classical Model and present their Emergent Model of forgiveness. This model conceives of forgiveness as a process of world-building between victims and perpetrators, which involves a complex combination of gestures that rebuild trust. It is a gradual process by which perpetrators and the wider community affected by world-shattering wrongs come to recognise the moral status of victims.¹⁸ Through this 'normative repair,' victims and perpetrators once again come to trust each other, thus enabling the wrongdoing to be forgiven.¹⁹

The notion of world-building as integral to forgiving world-shattering wrongs says something interesting about the nature of forgiveness. By adapting these insights into a more generalised form, a more detailed picture of the Relational Account may be fleshed out. This will be the goal of the following section.

VII. THE RELATIONAL ACCOUNT OF FORGIVENESS

In Section 5 of this paper, two questions regarding forgiveness were raised: firstly, what it is for a subject to be in an appropriate standing to forgive, and what function it uniquely performs. This paper proposes that an analysis of the Emergent Model yields answers to both questions.

The characterisation of forgiveness as a normative repair in relationships is instructive in this regard. It indicates two necessary conditions for forgiveness as a general phenomenon – first, there must exist some relationship against which forgiveness takes place. Secondly, there must be some damage in the relationship

¹⁶ *Ibid.*, 49.

¹⁷ Simon Wiesenthal, "The Sunflower: On the Possibilities and Limits of Forgiveness (New York: Schocken Books, 1998), 9.

¹⁸ Carse, Tirell, "Forgiving Grave Wrongs," 53.

¹⁹ It should be clarified that, by Carse and Tirell's admission, neither the recognition of a victim's moral status nor the rebuilding of trust are sufficient for forgiveness; Carse and Tirell argue that it is merely necessary for forgiveness to be possible in world-shattering wrongs.

in need of repairing. Because the repair is normative in nature, the damage in need of repairing must be a kind of wrongdoing. Thus, we may glean a promising answer to the question of what constitutes an appropriate standing to forgive – S1 is in an appropriate standing to forgive S2 only if S2 has wronged S1²⁰.

Our conclusion for the second question follows naturally from the first, and is self-evident from the characterisation of forgiveness within the Emergent Model – the function that forgiveness performs uniquely is that of normative repair in relationships. Put plainly, forgiveness is unique in restoring relationships that are damaged by some wrongdoing. However, this begets a subsequent question – how is it that forgiveness repairs relationships in this manner?

The Emergent Model highlights that, in the context of world-shattering wrongs, forgiveness repairs relationships through its gradual process of world-building for the victim. However, the metaphor of ‘world-building’ falls short when contextualised in a more general form of forgiveness. In comparatively more trivial cases of forgiveness, victims’ worlds are not shattered, and there is no world to rebuild – if someone lies to me, it is not often the case that my sense of self, worldview, or perception of reality is compromised – my ‘world’, as it stands, remains intact. I have simply and plainly been wronged, without my world being shattered. Yet, as was earlier established, my being wronged by someone who lied to me puts me in an appropriate standing to forgive them. How, then, does my forgiving them achieve the normative repair of our relationship?

This paper proposes that normative repair is achieved through a rebuilding of trust between the victim and the perpetrator. Trust is necessarily damaged after any experience of wrongdoing – if someone lies to me, my trust in them is plainly compromised. If a friend of mine does not show up to an arranged meeting, my trust in their ability to keep their promises to meet is compromised. In cases of world-shattering wrongs, trust is compromised on a spectacular level – a genocide eradicates victims’ trust that the perpetrators, and even the world at large, would afford them basic human dignity and rights.

This relational account of forgiveness provides further benefits in explaining how forgiveness can create relationships that were stronger than what they were before wrongdoing was committed. Consider the following case:

A and B are friends. A wrongs B. A makes amends, eventually resulting in forgiveness. Having reaffirmed their friendship, A and B find that they are closer than they had been before.

The case above is not an uncommon occurrence. Its permutations pervade our everyday lives, and are by no means out of the ordinary. Yet, other models may find it difficult to explain why it is that ‘post-forgiveness’ relationships may create stronger bonds. The Hamptonian Account would imply that the relationship is only restored to its original state, rather than improved. If forgiveness simply entailed that B comes to possess no moral hatred for A, it is not the case that B evaluates A as ‘more decent’ than before in forgiving them – B simply comes to

²⁰ ‘Wronging’ as used here admits of both direct and indirect wrongdoing. Thus, cases of indirect forgiveness would still count as cases of forgiveness – for example, a parent may forgive someone for murdering their son, because they were indirectly wronged by being robbed of their son. There is space for discussion on the limits of the indirectness of wronging, but that is not the interest of this paper.

adopt a similar assessment of the A's moral character as they did before A wronged them. As a general observation, Standard View accounts do not easily explain this phenomenon.

However, if it were the case that the function of forgiveness is normative repair, and this normative repair is achieved through a rebuilding of trust, then this phenomenon can be explained rather straightforwardly. Relationships may be strengthened after forgiveness because the parties in the relationship come to trust each other more. In the above case, A makes amends with B by rebuilding trust between them. In doing so, their trust is compounded – beyond simply trusting each other as much as before, they come to trust that they are capable of restoring their relationship after it is strained. From the restored relationship emerges a confidence that, come what may, A and B are capable of reconciling and once again seeing eye to eye. Thus, their trust goes deeper than what it was before; they trust each other to be decent human beings, but additionally trust that, if one does wrong, they are capable of repairing the resultant breach of trust. Taking this into account, it is difficult to imagine how normative repair may be said to occur through anything other than trust at all – in cases where relationships emerge from forgiveness stronger than they originally were, one would be hard pressed to characterise it in any terms other than trust.

This paper now consolidates its findings into a singular, united formalisation of what it is for some act to count as an act of forgiveness.

The earlier investigations into various accounts of forgiveness have yielded the following parameters that the formalisation must satisfy.

Standing: The analysis of Hampton's Account has revealed that, on pain of confusing social pardons with acts of forgiveness, any characterisation of forgiveness must exclude cases where one is not in an appropriate standing to forgive.

Non-Transaction: The analysis of the Emergent Model has revealed that, on pain of rendering the forgiveness of world-shattering wrongs impossible, any characterisation of forgiveness must not be merely transactional in nature.

These parameters, coupled with this section's argument for the centrality of trust in characterising forgiveness, yields this paper's suggestion for what is essentially involved in forgiving. Call this the Relational Criterion.

Relational Criterion:

What is essentially involved in S1 forgiving S2 is as follows:

- i) S2 had wronged S1, resulting in a damaged relationship, and
- ii) S1 and S2 repair their damaged relationship through a rebuilding of trust.²¹

²¹ In their Emergent Model, Carse and Tirell thought world-building involved perpetrators and wrongdoers, but also their wider society. This is because, in world-shattering wrongs, the community and world are also relevant stakeholders. In most cases of forgiveness, only perpetrators and wrongdoers are involved in the process of rebuilding trust. For this reason, the Relational Account omits community as a relevant stakeholder.

The Relational Criterion satisfies the requirement of Standing because condition i) of the Relational Criterion excludes cases where one is not in a standing to forgive. It also satisfies the requirement of non-Transaction because the process of rebuilding trust is non-transactional in nature.²² Condition ii) not only serves as a means to exclude one-sided attempts at forgiveness, but also by extension exclude transactional exchanges from being cases of forgiveness (at least, not by virtue of their being transactional).

This last point warrants further elaboration. Rather than being a unilateral process, or something that a victim simply bestows unto a wrongdoer, forgiveness is instead marked by a cooperative process of rebuilding the victim and wrongdoer's damaged relationship. In any other kind of relationship mending, this process must be two-way rather than unilateral. As a kind of relationship mending specifically aimed at repairing relationships damaged by wrongdoing, forgiveness is no exception. This feature qualifies the Relational Account's conception of forgiveness as being non-transactional.

Some clarification on the nature of forgiveness under the Relational Account may also be needed. Because of the centrality of trust in the account, and trust being a kind of attitude, one may be led to believe that the Relational Account is a kind of Standard View, characterising forgiveness as an attitude or change in attitude. This is not the case. While trust is inextricably linked to the concept of forgiveness, forgiveness does not refer to the restoration of trust itself, but rather the normative repair of a relationship that trust restoration is necessarily a part of. If trust is restored but a relationship is not repaired, then forgiveness is not obtained.

For these reasons, this paper claims that the Relational Account overcomes all of the challenges it has put forward to competing accounts, while also maintaining its identity as a distinct account in itself. It thus concludes that the Relational Account is, of the accounts put forward in the paper, the most accurate in characterising what it is for some act to count as an act of forgiveness.

This last point warrants further elaboration. Rather than being a unilateral process, or something that a victim simply bestows unto a wrongdoer, forgiveness is instead marked by a cooperative process of rebuilding the victim and wrongdoer's damaged relationship. In any other kind of relationship mending, this process must be two-way rather than unilateral. As a kind of relationship mending specifically aimed at repairing relationships damaged by wrongdoing, forgiveness is no exception. This feature qualifies the Relational Account's conception of forgiveness as being non-transactional.

Some clarification on the nature of forgiveness under the Relational Account may also be needed. Because of the centrality of trust in the account, and trust being a kind of attitude, one may be led to believe that the Relational Account is a kind of Standard View, characterising forgiveness as an attitude or change in attitude. This is not the case. While trust is inextricably linked to the concept of forgiveness, forgiveness does not refer to the restoration of trust itself, but rather the normative repair of a relationship that trust restoration is necessarily a part of. If trust is restored but a relationship is not repaired, then forgiveness is not obtained.

²² This is claimed using the same line of argument used to justify the Emergent Model as distinct from the Classical Model of forgiveness. For more details, see Section 6.1.

For these reasons, this paper claims that the Relational Account overcomes all of the challenges it has put forward to competing accounts, while also maintaining its identity as a distinct account in itself. It thus concludes that the Relational Account is, of the accounts put forward in the paper, the most accurate in characterising what it is for some act to count as an act of forgiveness.

VIII. OBJECTIONS

Two objections to the Relational Account will now be addressed. First, the paper considers the potential counterexamples that self-forgiveness and forgiving dead people may pose. Then, the paper considers a counterexample from a hypothetical person who has boundless trust in someone who wronged them.

Questions may be raised regarding how the Relational Account would treat unique cases of forgiveness, such as self-forgiveness, or the forgiving of dead people. If this objection is sound, it would be more likely that self-forgiveness or forgiving dead people were simply cases of forgiveness that the Relational Account fails to characterise as cases of forgiveness, than it would be that the two are not cases of forgiveness at all. In other words, if the Relational Account cannot characterise self-forgiveness or forgiving dead people as cases of forgiveness, it would be quite likely that the Relation Account is unnecessary for some act to count as an act of forgiveness.

The crux of this objection is that it is unclear how one might participate in a relationship with oneself, or with a dead person. Thus, it would seem that cases of self-forgiveness or forgiving dead people fail on condition i) of the Relational Criterion.

In order to resolve this objection, it is worth considering what kind of relationship is relevant in the Relational Criterion. While this paper acknowledges that there exists room for debate on whether relationships require two-way interaction (such that one cannot have a relationship with oneself or with dead people), it holds that, for the purposes of how the Relational Account treats relationships, one can in fact hold a relationship with oneself or with dead people. This is because the relationship described in the Relational Criterion is characterised by wrongdoing – one bears the relationship described in condition i) if they have wronged another or have been wronged by another. With this in mind, it becomes clear that one can in fact bear the relevant relationship to oneself, or to dead people – it is possible to wrong oneself or be wronged by oneself, and it is equally possible to wrong dead people or be wronged by dead people (though obtaining forgiveness from a dead person is obviously impossible).

Given that it is possible for some cases of self-forgiveness or forgiving dead people to fulfil condition i), what remains is to demonstrate that they may fulfil condition ii) – to resolve this objection, it must be shown that it is possible to rebuild trust with oneself, or with a dead person.

One may rebuild trust with oneself in a similar way to how one rebuilds trust with others – through small gestures that work towards reconciliation and amendment. If I have wronged myself through my self-destructive tendencies, then it is through my resolution of my self-destructive tendencies and restoration of my self-esteem that I may be said to have rebuilt my trust in myself (specifically, my trust that I will not engage in activities I know would harm myself).

While it is not the case that dead people may take action towards rebuilding trust when they are dead, it is still possible for someone to rebuild their trust in a

dead person after they have died. It might be that, after their death, one discovers some truth about the wrongdoing or the dead person's character that causes them to restore their trust in them. Alternatively, it may be that the deceased had taken action towards rebuilding trust prior to their death, and one only comes to be in the right emotional state to recognise these actions and restore their trust in them after their death. In both of these cases, condition ii) is fulfilled despite the wrongdoer being dead.

Thus, we observe that in both self-forgiveness and forgiving dead people, there are possible situations where the Relational Criterion is satisfied. It is hence possible under the Relational Account to forgive oneself, or to forgive dead people. It can therefore be concluded that self-forgiveness and the forgiving of dead people are not counterexamples to the Relational Account.

Next, this paper considers a counterargument from boundless trust. The counterargument may be explicated from the following case:

I have a boundless amount of trust in my friend – that is to say, there is nothing my friend can do that would make me trust them less. One day, I find out that my friend had taken some of my money without my consent – they had wronged me. However, because my trust in my friend is boundless, I do not think they have done anything that amounts to a breach of trust. Thus, despite being wronged, I maintain my high level of trust in my friend. My friend reaches out to rebuild the relationship. I willingly reciprocate, not believing any breach of trust has been committed but being eager to further our relationship.

My friend had wronged me, we had rebuilt our trust, and we had done so collaboratively. Thus, all the conditions for the Relational Account seem to have been satisfied, on the Relational Account's terms, it would seem that I had forgiven my friend. However, it seems obvious that I had not forgiven my friend in this case – this is simply my doing a disservice to myself through my own blindness or gullibility. The worry here, then, is that the Relational Account would wrongly classify certain cases as cases of forgiveness.

This counterargument may be resolved by clarifying that condition ii) had not been satisfied in the case of boundless trust. Repairing necessitates that something must have been damaged that needs to be repaired – one cannot repair that which is undamaged. Because my trust in my friend is boundless, my trust in my friend was not damaged from their stealing my money; it cannot be damaged at all. Thus, because my trust in my friend cannot be damaged, I cannot forgive my friend so long as my trust in them remains boundless, as there would be no circumstance under which my trust in them can be repaired, and thus no situation under which condition ii) of the Relational Account may be satisfied. As such, boundless trust cases would never be considered cases of forgiveness under the Relational Account, and the Relational Account does not wrongly characterise such cases as cases of forgiveness.

IX. CONCLUSION

The Relational Criterion is not intended to be both necessary and sufficient for an act to count as an act of forgiveness. This would be far too ambitious a task for one paper alone to accomplish. Furthermore, if the Relational Account were sound, there is reason to doubt that one may conceive of an account of forgiveness that necessarily and sufficiently characterises forgiveness in the first place. If forgiveness truly were fundamentally a kind of repairing done to a relationship, then it is likely that there are a plurality of means by which one may achieve forgiveness, and thus a plurality of iterations of forgiveness that vary from relationship to relationship. The fact that forgiveness is situated exclusively in the context of relationships damaged by wrongdoing, coupled with the truism that such relationships may take a multitude of forms, would imply that an account that is both necessary and sufficient may be impossible to achieve, or at least extremely difficult. After all, what may be sufficient for an act to count as an act of forgiveness in one relationship may not be sufficient in another. However, this does not entail that there does not exist certain unifying features that all forms of forgiveness have in common. It is the hope of this paper that the Relational Account is a step forward in identifying as many of these features as possible.

BIBLIOGRAPHY

Allais, Lucy. "Elective Forgiveness", *International Journal of Philosophical Studies* 21, no. 5 (May 2013): 637-653. <https://doi.org/10.1080/09672559.2013.767525>

Blustein, Jeffrey. *Forgiveness and Remembrance: Remembering Wrongdoing in Personal and Public Life*. Oxford: Oxford University Press, 2014.

Carse, Alisa L.; Tirrell, Lynne. "Forgiving Grave Wrongs", in *Forgiveness in Perspective*, ed. Christopher Allers & Marieke Smit (Amsterdam: Rodopi Press, 2010), 43-65.

Heyd, David. "Supererogation", *The Stanford Encyclopedia of Philosophy*, Winter 2019, Edward N. Zalta.

Fricke, Christel. *The Ethics of Forgiveness: A Collection of Essays*. Oxford: Routledge, 2011.

Griswold, Charles. *Forgiveness: A Philosophical Exploration*. Cambridge: Cambridge University Press, 2007.

Hampton, Jean. *Forgiveness and Mercy*. Cambridge: Cambridge University Press, 1988.

Hieronymi, Pamela. "Articulating an Uncompromising Forgiveness". *Philosophy and Phenomenological Research* 62, no. 3 (2001): 529-555.

Holmgren, Margaret. *Forgiveness and Retribution*. Cambridge: Cambridge University Press, 2012.

Kolnai, Aurel. Forgiveness. *Proceedings of the Aristotelian Society* 91, no. 1 (1973): 91-106. Lauritzen, Paul. Forgiveness: Moral Prerogative or Religious Duty? *The Journal of Religious Ethics* 15, no. 2 (1987): 141-154.

Murphy, Jeffrie. *Getting Even: Forgiveness and its Limits*. Oxford: Oxford University Press, 2003.

Norlock, Katherine. *Forgiveness from a Feminist Perspective*. Minneapolis: Lexington Books, 2009.

Pettigrove, Glen. *Fitting Attitudes and Forgiveness*. Oxford: Oxford University Press, 2021.

Thomason, Krista. Forgiveness or Fairness? *Philosophical Papers* 44, no. 2 (2015): 233-260.

Warmke, Brandon. *Forgiveness and its Moral Dimensions* Oxford: Oxford University Press, 2021.

Wiesenthal, Simon. *The Sunflower: On the Possibilities and Limits of Forgiveness*. New York: Schocken Books, 1998.

A Phenomenological Experience of the Novel

Virginia Moscetti
Swarthmore College

FOREWORD

Editors Ben Knepper, Alejandro Maris Natera

We have all felt the riveting sensation of experiencing a novel as “real.” As readers, we gain a sense of meaning in the novels we read. The philosophical underpinnings of these levels of meaning are precisely what Moschetti elucidates in “A Phenomenological Experience of the Novel.” Through actively *doing* phenomenology, Moschetti challenges the common belief that the realm of fiction is somehow distinct from that of reality. Moschetti skillfully demonstrates how key tenets of Husserlian phenomenology in *Ideas I* and *Ideas II*, including “givenness,” “at hand,” “horizons,” and “intersubjective community,” clarify three ways in which we find meaning and reality in novels: First through our relation to the novel as a physical object in our surrounding environment, second through our interpretation of the fictional world created by the novel, and lastly through our experience of the concrete details of the novel. Moschetti reveals that the latter literally intertwines the spatial-temporal worlds of the novel and its reader, allowing the reader to comprehend and foreshadow the novel’s development.

By showing how this explanation is significant to the narrative structure, Moschetti opens a new perspective on how we read a novel by imagining ourselves in the space of its details instead of a passive, outside spectator. By an act as simple as turning pages, we plunge deeper into the fictional and while solidifying the real. If you enjoy reading, join Moschetti in the intersubjective community of fellow readers and experience a paper that will make you love reading even more.

ABSTRACT

The following paper evaluates the ways in which a novel can mean to its reader as an object and as a fictional world from a phenomenological standpoint through Edmund Husserl’s *Ideas I* and *Ideas II*. Paying close attention to concrete fictional details, this paper demonstrates how these details may imply narrative possibilities to the reader in the instance that they are encountered by invoking the structures of real experience. While structural literary analysis has historically deemed these details narratively insignificant, this paper attempts to illustrate how consulting Husserl’s phenomenology can help us better understand their narrative function.

For Edmund Husserl, the given world in general experience is that which is “on-hand” for me (or available for me to turn my attention towards) at any given moment both determinately and indeterminately. The world of the novel can similarly be understood as a given world. Like the given-world in general experience, it is on-hand for me to interact with in various ways, depending on how I approach it and what the author, through their descriptive, narrative, and other such choices makes accessible or “given” to me. As a result of how I interact with the totality of what is “on-hand” for me within the novel’s given world, that novel will mean to me as that novel in a particular way.

Concrete descriptive details such as tables, boxes, pianos, flowers (which I will here term “details” generally) are included in this totality. As objects which are on-hand for me within a given (albeit fictional) world, I encounter them as spatially and temporally located and can exert my expectations over them. While details are commonly viewed as significant in terms of their overarching aesthetic (i.e., how they are beautiful or create an experience of beauty) or narrative function (how they

serve the progression of plot) in structural literary analysis, from a phenomenological standpoint, their capacity to structure my experience and meaning-making of the novel (depending on the ways in which I attend to them) as things which are on-hand for me, is critically significant. As a result, details which do not seem to serve an explicit narrative or aesthetic function become phenomenologically significant as objects which are on-hand for me, or any reader, to turn towards and meaningfully attend to.

In this paper, by referencing Husserl's *Ideas I* and *Ideas II*, I will demonstrate the different levels through which a novel can mean to a reader, beginning first as a physical object which means to the reader as a book, progressing to a novelistic object which means to the reader as a fictional world, and examining finally how the reader's experience of concrete details structures their experience of the novel's given world and can suggest to them various narrative possibilities in the instance in which these details are encountered, regardless of whether or not these possibilities are ultimately realized within the novel.

While structural literary analysis tends to deem significant those details that indicate possibilities that have been narratively or aesthetically realized throughout the course of the novel, those details could not have been understood as significant if the reader had not first experienced them as suggesting some kind of narrative or aesthetic possibility. Examining how we experience novels, their fictionality, and the details they contain, will allow us to interrogate

(1) why novels mean to us as novels, (2) how we make meaning out of our encounters with details, and (3) the implications these may have for the way we conduct literary analysis.

When I pick up a novel, I can attend to it in any number of ways, whether that is by examining the cover, listening to the pages flutter, smelling the pages, flipping the book upside down, or throwing it against the wall. I can also decide to open the novel and begin reading in whichever way I choose. I can skim, I can analyze, I can immerse myself in the plot, etc. In all of these experiences, I "constitute, by means of the apprehensions [or sensory mechanisms] allotted to [me], corresponding features of the thing as such by way of adumbration", meaning that as I begin to construct the object as it means to me as that object through the perceptual adumbrations, I encounter in my experiences with it.¹ Looking at the novel, I note pages, its covers, its colors, the author's name, the book's title, the word novel printed somewhere and recognize that it means to me as a novel and that all of its constituent parts mean to me as those parts. Attending to the novel in these ways, it is determinately on hand for me. Still, I recognize that I am sitting at a desk with a box of tissues and a lamp resting on top of it. This desk, tissues, and lamp "without being themselves perceived, indeed, without being intuitively present themselves, [are] actual objects [which] are there for me as determinate, more or less familiar objects, in unison with the objects currently perceived".² Here, Husserl demonstrates that objects, while I may not be directing my attention towards them,

¹ Husserl, Edmund. *Ideas Pertaining To a Pure Phenomenology and Phenomenological Philosophy Second Book: Studies in The Phenomenology of Constitution*. Translated by Richard Rojcewicz and Andre Schuwer. III. 5th ed. Vol. III. III vols. Dordrecht/Boston/London: Kluwer Academic Publishers, 2000 (page 62). (Hereafter referred to as "Husserl, *Ideas II*" and page number)

² Husserl, Edmund. *Ideas for a Pure Phenomenology and Phenomenological Philosophy, First Book: General Introduction to Pure Phenomenology*. Translated by Daniel O Dahlstrom. Indianapolis/Cambridge: Hackett Publishing Company, 2014 (page 49).

are determinately on-hand for me insofar as they are co-present with the objects (in this case, the novel) that I am attending to. The novel, as a book-object, can never exist for me in spatial isolation, but always against a backdrop of other objects that are determinately co-present with it.

The novel along with its backdrop of other objects, which are determinately on-hand for me, are also “in part-pervaded, in-part surrounded by a horizon of indeterminate actuality.”³ This “horizon” refers to the broad wealth of expectations and connotations that I associate with a particular object. My expectations for an object pervade my experiences with it, not in a determinate way that can be captured by my consciousness (unless I am literally thinking about how I think that a particular object behaves) but in an indeterminate way that is related to my previous experiences with that object and other similar objects. Whenever I turn towards the novel, my expectations for the way that it looks, smells, behaves, can be moved, etc. can either be confirmed or denied. If confirmed, then the novel continues to mean to me as my idea of that novel. If denied, then I re-evaluate my expectations for that novel and, subsequently, the way that it means to me as that object.

Whenever I turn towards the novel, my expectations for it are motivated by the process of literally “turning-towards” it. If the novel continues to look and behave in a way that I would expect it to throughout my multiple and various experiences with it, then my expectations confirm themselves continuously and my constituted idea of that novel undergoes a kind of sedimentation in which I consolidate my expectations for it. Husserl terms this process of bodily or mentally “turning towards” or “attending to” which motivates my perceptual expectations for a certain object in order for it to continue to mean to me as that object, kinesthetic sensation.⁴ Once I constitute a “novel” by virtue of my perceptual, active (in which I might open it, close it, or throw it), reflective, and imaginative experiences (in which I represent the fiction in my head and engage with it) with the novel, the novel is “founded” for me. In other words, it now means to me as a novel. When I look at other novels, that meaning becomes given in different ways through my perceptions of them.

Importantly, when I read a novel that I understand to be fictional, I expect that the characters and their circumstances do not exist. Therefore, the novel means to me as fictional (i.e., I constitute it as fictional). Meaning to me as fictional, the world of the novel also means to me as distinct (in terms of actuality) from the world I can experience by acting in it tactically, audibly, visually, etc. That being said, when I read, I attempt to represent the world of the novel imaginatively within my head, and in so doing, immerse myself within my conception of the novel world. This immersion is informed by how the author explicitly presents it to me and by how the characters interact with their world spatially, temporally, perceptually, as well as in terms of values, desires, etc. As a result, when I attend to the novel, two distinct worlds become co-present and determinately on-hand for me: the given world of the novel which I represent imaginatively in my head and the world of general experience in which I can act. When I read, I attend to both with varying degrees of attentiveness. However, whether I am skimming absently through the novel and reaching for a tissue or fully engaged in the novel’s given

³ Husserl *Ideas I*, page 49

⁴ Husserl *Ideas II*, page 61

world, neither one of the “worlds” entirely disappears for me. Instead, they both remain on hand for me insofar as the novel is something which I am reading, otherwise engaging with, or something that is determinately co-present with the objects, things, thoughts that I might be presently attending to.

The world of the novel consists, in part, of concrete details. The way that these fictional details mean to me is framed by my own experiences with the actual objects they denote as well as how the author presents them to me. Even if I have never encountered the object denoted by a fictional detail, my experiences as a person in the world imply expectations for how that object might behave. For example, if an author mentions an object “uhgurhg”, before the author provides any further descriptive information, I will likely assume that “uhgurhg” occupies space and has a certain dimensional and geometric configuration. As a result, my experiences in the world outside the novel manifest themselves in my expectations of the details presented to me by the author. In turn, the author’s description of the details informs my conception of them, and consequent expectations for them, within the context of the novel. Therefore, when I encounter details, I bring in to play two distinct given worlds.

In his novel, *A Simple Soul*, Gustave Flaubert describes a barometer in relation to a pile of boxes and a piano: “on an old piano, under a barometer, there was a pyramid of boxes and cartons.”⁵ From this description, I can draw multiple conclusions.

First, this description orients the barometer in space. Interposing myself imaginatively in the novel’s world, I can gauge my spatial relationship to the barometer if I were standing by the piano. This imaginative activity is made possible by my experience in the world with other objects, in which my body occupies a spatial position in relation to those objects. For Husserl, “the body is the medium of all perception...it is involved [in experience] as a freely moved totality of sense organs. All that is thingly-real in the surrounding world of the Ego has its relation to the Body...The body [is] the bearer of the zero point of orientation.”⁶ Because the body comprises the totality of my normal sensibility (i.e., all of my senses are located within the body), it is the site upon which all sensory perception occurs. Through the body, I orient and move myself towards objects in the kinesthetic sense, motivating, through this process certain perceptual expectations for the objects I turn towards. As the site of sensibility and the vehicle through which I transport myself in space, everything in my experience stands in relation to it. I see an apple through my eyes. Seeing it through my eyes, my visual experience of the apple is in relation to my eyes and their relative ability to perceive the apple. I also see the apple as something which is not me, but distinct from me in that it is spatially distinct from the rest of my body. My body is the “ground zero” of spatial orientation; it is the location from which I experience everything that is not my body.

As a result, all objects stand for me in relation to my body. However, this applies not only to objects which are actually present for me in the world, but to objects which could possibly be present. Returning to the barometer, while I imagine it as occupying a certain spatial orientation in relation to the other objects described, I conceive it in relation to my body and sense organs in a particular way. This might be as “in front of me,” “to the right of me,” “to the left of me,” or etc. Each of these spatial orientations then imply different perceptual possibilities. For example, if

⁵ Flaubert, Gustave, Dagny, John Bickers, and David Widger. *A Simple Soul. The Project Gutenberg E-book of A Simple Soul*. The Project Gutenberg, n.d., Chapter 1, <https://www.gutenberg.org/files/1253/1253-h/1253-h.htm>

⁶ Husserl *Ideas II*, page 61.

I imagine the barometer in front of me, I expect to be able to experience a front-facing adumbration of the barometer. If I imagine myself to be sitting at the piano, I expect to be able to see the underside of the barometer.

Conceptualizing the fictional barometer in terms of how I, the reader, would experience it as something actually present in the world has important implications for our relationship with fictionality as readers. At first glance, interposing myself imaginatively within the room and conceptualizing the barometer in terms of how I might experience it if I were in that room seems to suggest a kind of self-abnegation in which I wholly immerse myself within the novel's given world, abandoning, in the process, the world outside of the novel.

However, my conception of the fictional barometer is structured by the same systems of experience that are embedded in my encounters with the "thingly-real" in the outside world. In the same way that I would experience the barometer in a particular spatial location in relation to my body in the outside world, so I conceptualize the barometer imaginatively in terms of a spatial location in relation to my body. Bringing in the way that I experience spatiality in terms of the "thingly-real" to my encounter with fictional details, my experience of spatiality in the world outside of the novel becomes manifest in my experience of the novel's world. This transposition of my experience of spatiality from the outside world to the novel world blurs the distinction between them in a way that critically compromises the novel's fictionality. Because I experience fictional details imaginatively in terms of how I experience actual objects tactically, visually, etc., I apply a certain degree of spatial actuality to the novel's details in my experience of them. By doing this, I diminish their fictionality from things which do not exist for me as actual objects to things which I can relate to as actual objects in my imagination. And yet, interestingly, this experience of spatial actuality with respect to the barometer is made possible by the level of particularity invested in its description. While I would likely attribute a spatial, geometric, and dimensional configuration to the barometer the instance that I encountered it in the novel, the way that it is described enables me to locate my body (thus invoking my experience of the general world) with respect to it. But, of course, the "way" that it is described is nothing more than a particular fictional construction. This descriptive, fictional "particularity", then, paradoxically makes possible an experience of the fictional in terms of how we experience objects in the real world. While fictionality, by invoking the structures of real experience, leads us to relate to its fictional objects as actual within our imagination (thereby compromising fictionality itself), because those objects could not be experienced as actual without the particularity fiction engenders, it simultaneously immerses us deeper into the fictional—insofar as we continue to transpose our bodies imaginatively within the un-real spatiality presented to us in the novel.⁷

⁷ In her critical work "The Rise of Fictionality", Catherine Gallagher elaborates upon the paradox of fictionality in novels. Essentially, she claims that fiction as a concept and category is distinct from fantasy, which does not make any referential claims towards real-experience that make it believable. In the case described above, the particularity with which Flaubert describes his barometer makes a reference to how we encounter objects in the real world (i.e., within and against a background of other objects that possess spatial relationships to each other), thus making it believable—or possibly encounterable in real experience—to the reader. In her work, Gallagher focuses on realist Victorian novels (which she claims inaugurated, through their emergence, our modern conception of fiction). Science fiction and fantasy novels, while their content is decidedly less realistic than George Eliot's *Middlemarch*, similarly employ techniques and conventions towards producing a referential sort of believability. (Gallagher, Catherine. "The Rise of Fictionality." Essay. In *The Novel Volume I: History, Geography, and Culture*, edited by Franco Moretti, 336–63. Princeton and Oxford: Princeton University Press, 2006.)

Secondly, Flaubert's description also orients the barometer temporally within a specific instance in the beginning of the novel (this can also be understood spatially since the "beginning of a novel" is something that I can physically return to by flipping back to the beginning of the book). Therefore, the space and time in which it is located helps me orient myself within the novel's fictional world. By being present in a spatial location (the room in which it is above the piano and cardboard boxes) and a temporal location (the beginning of the novel), I constitute it as something which is both in the room and in the beginning of the novel. It is through the constituting activity of consciousness that a spatiality and temporality is constituted in distinction from the constituting activity of perceiving consciousness (presumably in a background absorption that is only peripheral given my absorption in the world of the novel).

Thus, when I encounter the barometer, its spatiality and temporality become given in it as part of its "constitution" for me. By extension, in connoting a spatio-temporal location to me (the reader) the barometer helps orient the novel's fictional world in a sequential narrative progression in a way that can be understood and followed by the reader. This is to say that, in being spatio-temporally situated, each detail implies a specific space and time that are then followed by other specific spaces and times implied by other details which both pertain to and assert a specific spatio-temporality. For example, if after the description of the barometer in the room, a singular bee in a rose garden is described, we as readers note a change in space and a kind of temporal succession from image to image or description to description. Therefore, details can be understood as marking the progression of the narrative structure from space to space and moment to moment. More importantly, however, the barometer's spatial-temporal location implies certain possibilities (or the lack thereof) for the narrative. Located in that room in that specific time-frame in the novel, the barometer can be encountered by any character who enters that room at that time in the novel, regardless of whether or not there are any that actually do.

The fact that concrete details, by virtue of their spatio-temporality, suggest what can and cannot happen, what is possible and impossible in terms of what is on-hand for certain characters, to the reader attending to them, implies that they frame what is possible for the narrative to achieve. In this way, concrete details, to some extent, control the narrative.

While no explicit description of the barometer as looking, acting in, or doing something in a particular way is provided, I reference my previous experiences with, and knowledge of, a barometer in order for it to mean to me as a barometer. Knowing, then, that a barometer is an instrument used to measure atmospheric pressure used especially in forecasting the weather and determining altitude, that object means to me as something which can be used in a particular way to achieve a particular thing and, through that meaning, is also founded for me (i.e., the instance that I perceive a barometer when reading, its meaning as my conception of a barometer becomes given in it). My own knowledge and meaning of that specific barometer is also supplemented by the way that other characters interact with it. This is because I assume that a fictional character (created to imitate a "real-person" through methods of displaying interiority that I will not describe here), operates in the novel similarly to how a person might operate in a specific context within the real world.

According to Husserl, all ego-subjects, or persons, by virtue of living in, interacting with, and having on-hand, the same spatio-temporal world, exist as an “intersubjective community.”⁸ While all ego-subjects will have different experiences with the objects that are on-hand for them (depending on the particularities of their psychological makeup and the manner in which they attend to those objects), we expect that the way that an object means to us will harmonize with the way that an object means to others. If my expectations of a particular object (and thus the way that it means to me) do not correlate with those of another member of the intersubjective community, then I will either question my own expectation or question that of my fellow

ego-subject. For example, if I understand that a barometer is used to calculate atmospheric pressure and a fellow ego subject claims that it is used to calculate temperature, then our expectations of the barometer do not harmonize and one (or both) of us must re-evaluate and reform our expectations. As a result, what is understood to be objectively true of the barometer is that which is true for both myself and my fellow ego-subjects as well as everyone else in the intersubjective community.

In addition, because the barometer can be used similarly for similar purposes by other ego-subjects, it has an intersubjective use-value: as it means to me as something which can be used to measure atmospheric pressure, so it means to others (who have “founded” it perceptually or otherwise). If a character were to express a love for the barometer or proclaim it as especially valuable, that love and valuing becomes superimposed onto how that barometer means to me as a barometer which I am reading about in Flaubert’s novel. I now might experience the barometer as useful in a particular way—and thus potentially useful for the characters in that particular way—but also valuable to a particular character. Both of these meanings then become “given” in my subsequent experiences with the barometer.

Concrete details, understood as objects with intersubjective use-values, are also significant to the narrative structure. When I encounter Flaubert’s barometer as the reader, I constitute it in terms of the way that Flaubert describes it but also in terms of the way that I have previously experienced barometers. Because Flaubert does not describe the barometer in terms of sensual perceptions (like color or visual detail), I do not constitute his barometer in my imagination according to any perceptual information given to me. However, because I have previously encountered barometers within my own experiences, I constitute Flaubert’s barometer in accordance with my own past perceptual experiences (i.e., as being colored in some way and as having a particular form). Additionally, my past experiences with barometers denote that the barometer can be used for something, both by myself and by my other fellow ego subjects. What the barometer can be used for is therefore given (as a feature of my constituted image of a barometer) in how I encounter it in any context, including the fictional context of the novel. In the exact instance that I constitute the barometer as an object which does something and can be used for something, I formulate an expectation of it. Within the novel, this expectation presupposes that the characters (which imitate my own personhood but cannot lay claim to the embodied actuality that I have as a real person) use it in the same way. In other words, when I expect

⁸ Husserl *Ideas II*, page 89.

something of Flaubert's barometer, I expect it to be the same object which means to me as a barometer. Because I expect the characters' experiences of the barometer to harmonize with my own, if the characters use it to measure temperature or cradle and kiss it, my expectations for the barometer have been denied. At this point, I can either re-evaluate the representation of the characters as real people, re-evaluate my own expectation (and therefore constitution) of a barometer, or question Flaubert's representation of a barometer. Still, before my expectation is either confirmed (either directly or by omission if the barometer is never referenced again) or denied, my expectation that the barometer can be used by the characters in a particular way implies narrative possibilities. If the reader expects that the barometer can be used in such and such way, then narrative possibilities open up for them in which the barometer is used both in an expectation-confirming and expectation-denying way. While, most probably, the narration will gloss over the barometer, because the expectation occurred, narrative possibilities become available to the reader in the instance that they encounter the barometer. In sum, because the barometer has an intersubjective use-value, the reader can formulate expectations about how the characters will engage and encounter that barometer, which suggests possibilities concerning how that barometer might or might not be used. As such, the barometer, and other such concrete details, create narrative possibilities for the reader in the instance that the reader encounters them.

By contrast, in his critical work *The Reality Effect*, Roland Barthes claims that Flaubert's barometer has no narrative function because it does not "constitute an indication of characterization or atmosphere, and so can[not] finally be salvaged as part of the novel's structure."⁹ As Barthes argues, while the barometer (and other such concrete details) does not serve the purpose of characterization or atmosphere, it signifies the "category of the "real", and not its various contents... [and] becomes the true signifier of [formal] realism."¹⁰ For Barthes, this is because the novelistic concrete detail is a fictional referent to the real rather than a real description of a feature of the real itself. Semiotically speaking, this means that the referent replaces the signified from the sign, making the concrete detail "a direct collusion of a referent and a signifier", or a conglomeration of referent and a linguistic pattern associated with that referent (and, of course, that feature of reality that the referent points to).¹¹ Because the referent, being fictional, is not describing some feature of reality as it is in actuality, the sign itself (constituted by the signifier and the signified) cannot exist. For example, with respect to an actual barometer, the word operates as a signifier of the concept, or image of the barometer (the signified) and vice versa. However, with respect to Flaubert's fictional barometer, the word operates as a signifier of the referent, thus signaling a kind of pointing towards the external category of reality as a whole rather than describing some feature of that reality itself. Not actually designating a real, non-fictional barometer, "the signified is expelled from the sign [i.e., the combination of signifier and signified], and along with it, of course, there is eliminated the

⁹ Barthes, Roland. "The Reality Effect." Essay. In *French Literary Theory Today: A Reader*, edited by Tzvetan Todorov, translated by R. Carter, 11–16. Cambridge u.a.: Cambridge Univ. Pr. u.a., 1982, (page 11). (Hereafter referred to as Barthes and page number).

¹⁰ Barthes, 16.

¹¹ Barthes, 16.

possibility of developing a form of the signified, that is, the narrative structure itself.”¹² In other words, because the referent does not point to anything in particular but rather references some general category of reality, the barometer has no and is not a “sign”; it has no semiotic form to give it meaning. Within the novel, this means that the concrete detail, because it references an external reality, (in which it does not itself exist, thereby making it impossible for that detail to constitute a “signified”) and yet has no narrative bearing within the novel that might give it narrative significance or meaning, has no “sign-structure”—it does not occur within a framework that allows it to designate or signify anything in particular within or without the novel except for its referencing a general category of the real.

However, it is precisely because the concrete detail is referential that makes it possible for it to imply narrative possibilities to the reader. Because the barometer invokes the category of the real, it invokes how we experience real objects, thereby implying narrative possibilities along the lines of inter-subjective use-values and how we experience space and time in the narrative. In fact, we can justify the narrative function of concrete details from a phenomenological standpoint through Barthes’ definition of narrative. For Barthes, “the general structure of the narrative...appears essentially predicative; to be extremely schematic, ignoring the numerous digressions, delays, changes of direction, or surprises which the narrative conventions add to this schema, it can be said that at each juncture of the narrative syntagm, someone says to the hero (or to the reader; it does not matter which): if you act in this way, if you choose this alternative, then this is what will happen.”¹³ Because the barometer references a general category of the real and invokes how we experience that real, it (1) allows us to interpose our bodies within a particular narrative space and time, (2) assume possibilities for characters to act in a particular way (i.e., points out and structures predictive “junctures” within the narrative), and (3) marks the progression from narrative juncture to juncture (by signaling changes of environment) as we move along the course of the novel’s plot. Thus, the referent, by invoking how we experience the real, retains its narrative significance for us and continues to exist within a sign framework in which the signifier promotes a possible way for the referent to mean to us (the readers) within the narrative.

Ultimately, then, a novel can mean to its reader in various ways depending on the manner in which the reader attends to it. Among others, a novel means to its reader as an object, as a fiction, and as a given world comprising details, characters, plot, and narration. In each of these “meanings”, our experiences within the outside world become manifest, informing our relationship to the novel-object, the fiction, the equipmental totality, including concrete details, of its given world. As our outside-world experiences become manifest in our novel reading, narrative possibilities crop up before us. The barometer can be used by character X to measure altitude because I can use it as such. The barometer in this particular room at this particular moment marks this particular narrative juncture which is transcended by a new description of a new concrete detail, say, a rose in a garden. Thus, the barometer, while it references real experience, continues to acquire narrative meaning through that reference, allowing it to act as a sign within the novel.

¹² Barthes, 16.

¹³ Barthes, 12.

BIBLIOGRAPHY

Flaubert, Gustave, Dagny, John Bickers, and David Widger. *A Simple Soul. The Project Gutenberg E-book of A Simple Soul*. The Project Gutenberg., n.d., Ch. 1. <https://www.gutenberg.org/files/1253/1253-h/1253-h.htm>

Husserl, Edmund. *Ideas Pertaining To a Pure Phenomenology and Phenomenological Philosophy Second Book: Studies in The Phenomenology of Constitution*. Translated by Richard Rojcewicz and Andrew Schuwer., 5th ed. vol 3. Dordrecht/Boston/London: Kluwer Academic Publishers, 2000.

Husserl, Edmund. *Ideas for a Pure Phenomenology and Phenomenological Philosophy, First Book: General Introduction to Pure Phenomenology*. Translated by Daniel O. Dahlstrom. Indianapolis/Cambridge: Hackett Publishing Company, 2014.

Gallagher, Catherine. "The Rise of Fictionality." Essay in *The Novel Volume I: History, Geography, and Culture*, edited by Franco Moretti. pp. 336-63. Princeton and Oxford: Princeton University Press, 2006.

Barthes, Roland. "The Reality Effect." Essay in *French Literature Theory Today: A Reader*, ed. by Tzvetan Todorov, translated by R. Carter, 11-16. Cambridge: Cambridge University Press, 1982.

How Wide is the Mind?
The Extended Mind,
the Narrow Mind, and
Critical Disability Theory

Kexuan Liu
University of Virginia

FOREWORD

Editors Ethan Kohn, Alkis Boukas

Beyond an intuitive belief, what properties make one believe their mind consists of their brain? Are the properties of the mind exclusive to the brain? In “How Wide is the Mind?” Liu outlines Andy Clark and David Chalmers’ Extended Mind theory, as well as objections to it proposed by Brie Gertler. The Extended Mind theory states that objects outside of our brain conserve the same role as our brains, and should therefore be considered part of our mind.

Expanding on Clark and Chalmers’ original idea, this paper follows Otto, a man suffering from Alzheimer’s disease, and his reliance on a notebook that he keeps with him at all times. After defending Clark and Chalmers’ work from Gertler’s critiques, Liu applies the Extended Mind theory to challenge our societal views about disabilities, positing that they are societal constructs instead of inflexible biological realities.

This paper’s greatest strengths lie where Liu provides concrete, practical examples of how an extended mind must operate. In particular, the discussion of extended mind as it relates to critical disability theory is not only a straightforward and compelling application of this rather abstract philosophical concept, but also presents a fascinating new view that should enhance our understanding of cognitive disabilities. In doing so, Liu has not only contributed to our understanding of the extended mind in the context of Philosophy of Mind, but also in the context of Cognitive Science. This excellent work demonstrates that even a theory as initially unintuitive as extended mind may be elucidated by a well-reasoned and well-expressed argument.

ABSTRACT

The extended mind theory claims that items external to the body can contain our mental states, while the narrow mind theory argues that only occurrent states are part of the mind. In this paper, I will defend Clark and Chalmers’s extended mind theory in face of Brie Gertler’s most powerful objection (i.e., the proliferation of actions objection) and her alternative theory (i.e., the narrow mind theory). I will argue first that the (unwanted) consequence of proliferation of actions is also implied by Gertler’s narrow mind theory; and second that this consequence, far from being unwanted, is rather a desirable result. I will also engage with critical disability theory to provide a positive argument for the extended mind theory. Specifically, I argue that the extended mind theory is appealing and revealing in that it challenges three traditional assumptions about disability, the mind, and the self; and it reveals that the deep entanglement between the brain and its environment is not merely contingent but rather an existential and ontological condition. In this way, I think the extended mind theory provides a more accurate model of the mind and its relationship with the environment.

I.

The extended mind theory argues that the mind extends into the world. By parity, if items outside of our skin have the same cognitive function as the mental states internal to our body, then those external items should also be considered as part of our mind. In other words, things outside of our skins and skulls, when

playing the right sorts of cognitive roles, become the extension of our minds.¹

For example, a person called Otto suffers from Alzheimer's disease, and he relies heavily on his notebook to navigate through his life. Whenever he learns new information, he writes it down in his notebook, and whenever he needs to perform certain tasks, such as going to the Museum of Modern Art (MoMA), he consults his notebook about this information. Since the notebook plays the role that is usually played by a biological memory—that is, storing Otto's standing or non-occurring beliefs (i.e., beliefs that he is not currently entertaining or conscious of)—C&C argue that the notebook is Otto's extended mind.

To be more specific, C&C claim that for something to be an extended mind, it needs to meet several conditions. We can take Otto's notebook as an example. First, beliefs stored in the notebook influence Otto's actions as much as beliefs stored in one's brain. Also, the notebook is a constant in Otto's life in a sense that Otto rarely takes action without consulting his notebook. Third, the notebook is available to Otto without difficulty. Fourth, Otto automatically endorses information in the notebook and also consciously endorsed it at some point in the past.² Therefore, Otto's notebook is his extended mind.

For clarity and convenience, I will reconstruct C&C's argument with distinct premises and conclusion, as follows:³

- (1) Suppose that in a normal case where beliefs are embedded in biological memory, a person—for example, Inga—has a non-occurring or standing belief (i.e., beliefs that she is not currently entertaining or conscious of) about the location of MoMA, and she walks to MoMA after consulting her memory.⁴
- (2) Suppose that a person with Alzheimer disease writes down the same information about the location of MoMA in his notebook, and he walks to MoMA after consulting his notebook.⁵
- (3) “What makes some information count as a belief is the role it plays.”⁶
- (4) In both cases the information plays the same role (from [1] and [2]).
- (5) By parity, information stored in Otto's notebook should be counted as his standing beliefs (from [3] and [4]).
- (6) Standing beliefs are part of the mind.⁷
- (7) Otto's notebook is an extension of Otto's mind (from [5] and [6]).
- (8) Otto's notebook is beyond his skin and skull (i.e., external to his skin).
- (9) The mind extends into the world beyond one's skin and skull (from [7] and [8]).

¹ Andy Clark, and David Chalmers, “The Extended Mind,” in *Philosophy of mind: Classical and Contemporary Readings*, ed. David Chalmers (New York: Oxford University Press, 2021), 508.

² Clark and Chalmers, “The Extended Mind,” 511.

³ Gertler also offers a very clear and helpful reconstruction of C&C's argument in her *Overextending the Mind*, see Gertler, “Overextending the Mind,” 514.

⁴ Clark and Chalmers, “The Extended Mind,” 508.

⁵ *Ibid.*, 508.

⁶ *Ibid.*, 509.

⁷ *Ibid.*, 511.

II.

In this second section, I will introduce Gertler's most powerful objection to the extended mind theory, which argues that the extended mind theory brings about unwanted consequences such as the proliferation of actions. I will also introduce her alternative theory—the narrow mind theory.

Gertler argues that the extended mind theory will bring about unwanted consequences, one of which is that minds would extend to what we usually do not want to count as minds and lead to a proliferation of actions. For example, we could imagine that Otto's standing beliefs, instead of being stored in a notebook, are stored in external computing devices that are plugged into humanoid robots owned by Otto.⁸ These robots would perform different tasks based on these standing beliefs, one of which might act according to Otto's beliefs that "I need to make a banana bread" and that "I should do 1,2,3,4,5 steps to make a banana bread," and bake a banana bread successfully. Another humanoid robot might act according to Otto's belief that "I need to go to APA to give a philosophy talk" and indeed does go to APA and give a talk. While one of the humanoid robots goes to APA and the other starts baking a banana bread, the organic body of Otto might still be asleep in bed. Importantly, these robots are linked together in some ways and are in constant communication with each other; also, their beliefs are readily accessible to Otto. Hence, Gertler asks, should we say these behaviors are all Otto's actions? According to C&C's extended mind theory, it seems that these are all Otto's actions, for we can say that these robots store Otto's standing beliefs that are readily accessible and automatically endorsed by Otto. If so, there is "no limit to the actions a single person can perform," and Otto would become extraordinarily active.⁹ This might seem like a weird intuition but seem to follow from C&C's view.

This extraordinary activeness of Otto would be an unwanted consequence, particularly because it implicates an unfair distribution of responsibility in terms of both rewards and punishments. For example, we might not want to say that Otto should take credit for things he does not seem to do (i.e., that he has baked the banana bread and gone to the APA conference while his organic body is still asleep). And in more general terms, we might not want to say that one can take credit for things one does not do, or that one is responsible for harms that one does not seem to cause.

It follows that in order to avoid this consequence, we need to object to one of C&C's premises that leads to the extended mind thesis. While some philosophers decide to object to premise (4), Gertler thinks that the most objectionable premise should be premise (6). That is, she thinks that if we do not want the consequence of the proliferation of actions, then we should not admit that non-occurrent beliefs stored external to our body (e.g., in the notebook) are parts of the mind.¹⁰ By parity, we also need to reject that the non-occurrent beliefs internal to our body (i.e., in our brain) are part of our minds. In this way, Gertler proposes an alternative

⁸ Brie Gertler, "Overextending the Mind," in *Philosophy of mind: Classical and Contemporary Readings*, ed. David Chalmers (New York: Oxford University Press, 2021), 516.

⁹ Gertler, "Overextending the Mind," 516.

¹⁰ See Gertler, "Overextending the Mind," 518-519, for a detailed explanation of why premise (6) is the most objectionable premise. To state it briefly, she thinks that objections to other premises of extended mind thesis can be defeated by simply modifying the notebook example.

theory called the narrow mind theory, which holds that “the mind is made up entirely of occurrent states and conscious processes. These include beliefs or desires that are now being entertained, conscious thoughts, emotions, and sensations, and conscious cognitive processes.”¹¹ Hence, Gertler accepts C&C’s more controversial claim that “when it comes to belief, there is nothing sacred about skull and skin,” but she draws from it a different moral—for she chooses to reject that non-occurrent beliefs or standing beliefs are also part of the mind.¹²

III.

In the third section, I will defend the extended mind theory against the proliferation of actions objection and attack the alternative narrow mind theory. My defense will contain two negative arguments for the extended mind theory. First, I argue that the proliferation of actions is not only implicated by the extended mind theory, but also by Gertler’s narrow mind theory.

Hence, if she thinks that proliferation of actions is an unwanted consequence, then the narrow mind theory should also be rejected on the same ground. Second, I argue that proliferation of actions is not an unwanted consequence but rather an appealing result.

My first argument is that Gertler’s narrow mind theory cannot avoid the consequence of proliferation of actions as well. Let us revisit Gertler’s revised example of Otto: Otto’s standing beliefs, such as “I need to bake banana breads” or “I need to go to grocery stores today,” are stored in computing devices that are then plugged into an enormous fleet of interconnected, humanoid robots. More importantly, these humanoid robots would act according to these beliefs.

Given this revised example, I think beliefs stored in the humanoid robots are in fact occurrent beliefs, since when robots are acting, these beliefs are, accordingly, beliefs that are now being entertained. Yet, these beliefs are falsely characterized as non-occurrent beliefs of Otto’s by Gertler. Of course, it can be objected that only the mental states or beliefs that are now being entertained by the internal, organic brain count as occurrent beliefs. But this move counters Gertler’s alliance with C&C’s claim that “there is nothing special, vis-à-vis a subject’s agency, about states internal to her organic body.”¹³ Hence, if we think that beliefs being currently entertained by the organic brain are occurrent beliefs, then on the same grounds, we should say that beliefs entertained by the humanoid robots’ brains are also occurrent beliefs. If these beliefs are occurrent beliefs, then even according to narrow mind theory, actions derived from these occurrent beliefs will be Otto’s actions. It follows that narrow mind theory cannot avoid the (unwanted) consequence of the proliferation of actions as well.

Gertler might object that (i) these robots’ behaviors are not caused by occurrent beliefs, because their “brains” are not conscious, and only beliefs that are currently entertained by the conscious mind are considered as occurrent beliefs; and that (ii) even if these behaviors are caused by occurrent beliefs, they are not caused by Otto’s occurrent beliefs. Hence, these robots’ actions would not be Otto’s actions according to the narrow mind theory.

¹¹ Gertler, “Overextending the Mind,” 520.

¹² *Ibid.*, 521.

¹³ *Ibid.*, 517.

I argue that both moves (i.e., objection [i] and [ii]) are begging the question. The question at issue here is precisely whether these computing devices in these humanoid robots should be considered as part of Otto's mind. If we assume that they are not part of Otto's mind, then certainly these beliefs stored in the computing devices would not be Otto's occurrent beliefs, and then these actions performed according to these beliefs would not be Otto's actions. But to assume this is to already assume the conclusion that we find controversial to begin with.

My second argument is that it is very plausible that we credit Otto and hold him responsible for the proliferate actions performed according to his multiple occurrent and seemingly non-occurrent beliefs.

Gertler seems to suggest that crediting Otto for making the banana bread and going to the philosophy conference, while Otto's organic body is still asleep in bed, sounds unfair or even ridiculous. By subscribing to the extended mind theory, we accredit Otto and hold him responsible for things he seemingly has not done. In my previous argument, I argue that this is also a consequence for the narrow mind theory. In my second argument, I will argue that it is not an unwanted consequence, but rather a plausible result.

I think the result of the proliferation of actions seems unpleasant at the first sight because we usually have a "narrow" conception of body and mind. But according to the account of the extended mind, both the computing devices and the organic brain are parts of Otto's mind, and it might follow that both the humanoid, robotic bodies, and the biological body, are parts of Otto's body. When we say Otto baked the banana bread, we might be saying that the organic part of Otto baked the banana bread, and we might also be saying that the non-organic part (i.e., the humanoid, robotic body) of Otto baked the banana bread. Similarly, when we say Otto is responsible for the failure of his philosophy talk, we are attributing responsibility to Otto, which includes both his organic part and his inorganic part. Given that beliefs and information stored in the computing devices are automatically endorsed by Otto, and that Otto rarely takes relevant action without consulting it, I think it is plausible to say that Otto (as a system consists of both humanoid brains and bodies and the organic brain and body) should be responsible for, and taking credit for the actions done by both the organic part and the non-organic parts of his body.¹⁴

Furthermore, I see the implication that "there seems no limit to the actions a single person can perform" as a very desirable one.¹⁵ I think a revealing insight provided by Gertler's revised example of Otto, is that in the extended mind theory (and in this case, also in the narrow mind theory), the mind can be decentralized and spread out into the world. It challenges the philosophical and cultural assumptions of how humans should look and act, and it challenges the traditional envision of an individualized, autonomous, non-fluid way of being-in-the-world, in which a person can only occupy a singular location at one time.

IV.

In this section, I will engage with critical disability theory to provide a positive argument for the extended mind theory. I argue that the extended mind theory is

¹⁴ Clark and Chalmers, "The Extended Mind," 511.

¹⁵ Gertler, "Overextending the Mind," 516.

appealing and revealing in that it challenges three traditional assumptions about disability, the mind, and the self, revealing that the deep entanglement between the brain and its environment is not merely contingent but an existential and ontological condition. Further, I think what C&C really reveal is that the boundary between the mind and the world is an arbitrary one, and an insistence of that strict boundary implicitly supports biological chauvinism and an ableist picture.

According to critical disability theory, both physical and mental disability should be considered as socio-culturally constructed conditions, instead of self-evident physical facts. There are numerous discussions on that. For example, Thomson writes, “stairs create a functional ‘impairment’ for wheelchair users that ramps do not. Deafness is not a disabling condition in a community that communicates by signing as well as speaking. People who cannot lift three hundred pounds are ‘able-bodied’ whereas those who cannot lift fifty pounds are disabled.”¹⁶ I suggest that we can make similar observations about the “disabled” or just minority minds.

Otto’s memory fails only when he has no external memory devices that help him navigate through life. People with dyslexia (i.e., reading difficulty) function perfectly well if they listen to audiobooks instead. And those who have terrible brain-bound spatial navigation might not get lost if they are assisted by Google Maps. I think these examples reveal two essential things.

Firstly, the line between disability and ableness is drawn culturally and also arbitrarily. Secondly, what counts as disability and what does not is always bound to the affordances of the current environment—namely whether the environment can afford people with what they need.

The extended mind theory is relevant to the critical disability theory in that it, similarly, attends to the very supplementarity of the brain to its environment.¹⁷ C&C write, “the brain develops in a way that complements the external structures, and learns to play its role within a unified, densely coupled system.”¹⁸ In Otto’s case, rather than pinning Otto down as having an incomplete, unstable mind, C&C attends to the way Otto’s internal brain coupled with things existing in external environments—the notebook, for example—that complement Otto’s biological memory and help Otto to live an easier life. On a shallow level, we can suggest that the extended mind theory points out how technologies or items outside our skin and skull play a huge role in our—and especially, mentally “disabled” or mental minorities’—lives. But more importantly, by claiming that our mind and cognitive processes extend into the world beyond our skin and skull, C&C radically challenge three traditional views: (i) the medical view of disability that views disability as a self-evident biological fact, (ii) the modernist view that mind or self is an enclosed, singular entity, and (iii) the neurocentric view that mind needs to be made of neurons.

The extended mind theory starts with Otto’s case, but it can move beyond Otto and provide new insights into the ontological condition of human beings as a whole. The interdependence between Otto and his notebook is not just an empirical and contingent condition—that is, Otto is not an exception, and his dependence on his external environment is not only due to him having

¹⁶ Rosemarie Garland-Thomson, *Freakery: Cultural Spectacles of the Extraordinary Body* (New York: New York University Press, 1996), 7.

¹⁷ Clark and Chalmers, “The Extended Mind,” 508.

¹⁸ *Ibid.*

Alzheimer's. Rather, this interdependence between the internal brain and the external environment can be seen as an existential or ontological condition of human beings in general, which implies that perhaps there is no absolute, intrinsic boundary between the world (i.e., environment) and the mind. In other words, the boundary itself is drawn arbitrarily.

Because of its arbitrariness, we should not stick to a neuro-centric view of the mind or subscribe to a type of biological chauvinism or ableist picture that regards reliance on non-organic matters as inherently inferior states, but should rather take the extendedness of mind as integral to the mind itself and treat the mind as ordinary in its interdependencies with its environment. In this way, we might gain a more accurate model of what the mind consists of, and what its relationship with the environment is.

V.

In conclusion, I defend C&C's argument that the mind extends into the world in this paper. I first object to Gertler's powerful objection— the proliferation of actions— by arguing that Gertler's narrow mind theory will lead to the same consequence, and that this consequence is rather a plausible result. Then, I provide a positive argument for C&C's extended mind theory with reference to discussions on critical disability theory. I argue that C&C's extended mind theory provides enormous insights into the socially constructed nature of the category of "disability"/ "ableness." It also provides a new insight into our ontological existence as consisting of the entanglements between mind and the environment. I think that by taking the extendedness of mind not as an exception but as integral to the mind itself, and by treating the mind as ordinary in its dependencies on the environment, we can have a more accurate model of what the mind consists of, and what its relationship with the environment is.

BIBLIOGRAPHY

Clark, Andy, and Chalmers, David. "The Extended Mind." In *Philosophy of mind: Classical and Contemporary Readings*, edited by David Chalmers, 505-513. New York: Oxford University Press, 2021.

Garland-Thomson, Rosemarie. *Freakery: Cultural Spectacles of the Extraordinary Body*. New York: New York University Press, 1996.

Gertler, Brie. "Overextending the Mind." In *Philosophy of mind: Classical and Contemporary Readings*, edited by David Chalmers, 513-522. New York: Oxford University Press, 2021.

Shildrick, Margrit. "Why should our bodies end at the skin?: Embodiment, boundaries, and somatechnics." *Hypatia* vol.30, no.1, 2015, pp. 13-29.

Split-Brains, Mental Unity, and the Combination Problem for Panpsychism

Robison Yost
Wright State University

FOREWORD

Editor Rachel Kodysb

This paper is an insightful dive into the phenomenon of consciousness. We often take our consciousness as a singular entity, a simple and essential awareness of life. However, the panpsychist view presents a skeptical challenge to this one-dimensional model of the mind. It creates space for discovering the potential breakdown of awareness into parts, combinations, or separate experiences. This paper does not offer a definitive view of consciousness, and in this lies the work's power. Yost plants doubt through a good understanding of core arguments against panpsychism and leaves the rest to the reader. The author lays out the framework of the mind instead. Yost defines experience and questions exactly where this experience is taking place.

Yost uses the work of David Chalmers, Bertrand Russel, Immanuel Kant, and Thomas Nagel to present complex existences of the mind in spaces outside of a unified brain. The author also presents an interesting perspective on split-brain research, which displays parts of patients' brains seemingly independently responding to stimuli. Split-brain provides some neurological basis to the philosophical theories of the aforementioned philosophers. Overall, this work gives one the feeling that they have unknowingly stepped into the sky. Intuition about the mind turns on its heels.

I. INTRODUCTION

In recent years, panpsychist theories of mind, which contend that consciousness is a fundamental property of matter, have received considerable attention from those seeking a solution to the mind-body problem. However, these theories are not without their problems. Perhaps the most substantial criticism of panpsychism is the *combination problem for panpsychism* (hereafter combination problem), which is roughly the assertion that experiences do not sum. In other words, panpsychism must explain how microexperiences come together in the brain to yield a unitary macroexperience.¹ This objection to panpsychism has intuitive appeal, but as will be demonstrated here, there are examples where intuitions may also run counter to it, sometimes in ways that are quite forceful. One such example is that of two-minded interpretations of split-brain research, which may confirm some assertions made by panpsychists. This article will apply this compelling interpretation of split-brain research to panpsychism in an attempt to inject skepticism into the combination problem.

II. ARRIVING AT PANPSYCHISM

Panpsychism arises as a solution to the major problems of the two dominant theories of mind: materialism and dualism. According to materialist theories of mind, all truths about mental events can be reduced to physical truths. On the other hand, dualist theories of mind argue that not all truths about mental events can be reduced to physical truths, and as a result, these mental events are not

¹ Phillip Goff, "Experiences Don't Sum," *Journal of Consciousness Studies* 13, no. 10-11 (2006). 53-61.

physical. Of course, both theories run into their own unique problems, but for materialism, the *knowledge argument* and the *modal argument* are particularly glaring.

The knowledge argument is most notably described by Frank Jackson in *Epiphenomenal Qualia*. Jackson proposed a thought experiment in which a scientist named Mary knows everything that can be known about the science of color but has spent her whole life in a black and white room and so has never actually *experienced* color.² If, when Mary finally leaves the room, she learns something new about color through phenomenal experience that she could not discern from her studies in the black and white room, then materialism is false.

The *modal argument* states that if it is possible for a brain to function without producing consciousness, then consciousness cannot be reduced to physical truths. This argument is commonly presented as a conceivability argument, namely the conceivability of *zombies*, which are hypothetical beings that are identical to humans in every way (including the brain) except that they lack consciousness or anything that it is like to be them. The argument is roughly that the conceivability of zombies implies their possibility, and if zombies are possible, then materialism is false.³

By contrast, dualism faces one rather significant problem regarding causation.

Typically, the physical world is thought to be causally closed, meaning that all physical events are effected only by other physical causes. However, suppose dualism is correct in its assertion that consciousness is nonphysical. In that case, consciousness must be somewhere “outside” of the physical world, from where it can exert no causal power upon it. In other words, it is an *epiphenomenon*. Epiphenomenalism is generally considered to be an unattractive position, but some philosophers have come to embrace it.⁴

Disagreements between materialist and dualist theories of mind are centralized around a primary issue, namely, the issue of grounding consciousness in matter. This, however, is the primary strength of panpsychism, which maintains that some rudimentary form of consciousness is a fundamental property of matter in its simplest form. Panpsychism then takes consciousness as primitive, and by doing so evades the intractable problems of both materialism and dualism. This has made it a tempting proposition for philosophers, and as a result, the theory has exploded in popularity. Now that it is clear how one could arrive at a panpsychist position, let us properly define it.

III. PANPSYCHISM

There are many flavors of panpsychism, but only a few successfully address our central grounding issue. In *Panpsychism and Panprotopsychism*, David Chalmers describes the current landscape of panpsychist theories and advocates

² Frank Jackson, “Epiphenomenal Qualia,” *Philosophical Quarterly* 32, no. 127 (1982).

³ “What it is like” to be something is a frequently used definition of phenomenal consciousness³ that originates from Thomas Nagel’s 1982 article, “What is it Like to Be a Bat?”. Nagel’s point is that if there is something *that it is like* to be a thing, then that thing is conscious; Chalmers, *The Conscious Mind: In Search of a Fundamental Theory*, 93-171.

⁴ Chalmers presents one of the most forceful arguments in favor of epiphenomenalism in *The Conscious Mind* (1996, p. 150-161).

for a form of panpsychism that is both *constitutive* and *Russellian*.⁵ According to Chalmers, “*constitutive panpsychism* is the thesis that macroexperience is (wholly or partially) grounded in microexperience.”⁶ By this view, the experiences of fundamental constituents of matter combine to yield macroexperience when correctly arranged, viz., a brain.

Russellian panpsychism takes Bertrand Russell’s view that “physics reveals the relational structure of matter but not its intrinsic nature.”⁷ This is to say that classical physics has a grounding problem of its own, describing a world that is exclusively built of relationships and devoid of qualities intrinsic to matter. The Russellian view of physics hypothesizes that microphysical entities do not simply *play roles* (mass, charge, spin) but that there are properties responsible for these roles; Chalmers calls these properties *quiddities*.⁸ If it is the case that there are quiddities (and it very well might not be), then according to Russellian panpsychism, some of them are phenomenal experiences.

Here, in constitutive Russellian panpsychism, we have a panpsychist theory that addresses phenomenal grounding. However, Chalmers notes that there is a second view that accomplishes this same goal, what he calls *constitutive Russellian panprotopsychism*.⁹ This is the view that some quiddities hold properties that are not necessarily phenomenal experience but come together to yield phenomenal experience when arranged correctly, viz., a brain. Chalmers notes that the disjunction of constitutive Russellian panpsychism and constitutive Russellian panprotopsychism is *Russellian Monism*.¹⁰ The distinction of whether some quiddities are phenomenal or proto-phenomenal is somewhat arbitrary for this discussion, and so Russellian monism is a satisfactory definition and is what will be meant when “panpsychism” is used hereafter.

IV. THE COMBINATION PROBLEM

Panpsychism faces several problems; the most obvious is that, although it seems to answer many controversial issues for the philosophy of mind, it appears to have no basis in observable fact. Thomas Nagel expresses this wonderfully in *The View from Nowhere*: “Though [panpsychism] has its attractions as a way of uniting the radically disparate elements that give rise to the mind-body problem, it also has the sickening odor of something put together in the metaphysical laboratory.”¹¹ Defenders of panpsychism often dismiss this criticism (and probably for good reason), but the sentiment is not lost; perhaps there is something to say about gut-level intuition, especially when it applies to theories about conscious experience. At any rate, this objection to panpsychism will not receive additional attention here. Instead, what seems a far more forcible objection is what is known as the *combination problem*.

⁵ David Chalmers, “Panpsychism and Panprotopsychism,” 10.

⁶ *Ibid.*, 9.

⁷ *Ibid.*

⁸ *Ibid.*

⁹ Chalmers, “Panpsychism and Panprotopsychism,” 14.

¹⁰ *Ibid.*, 17.

¹¹ Nagel, Thomas. (1989). *The View From Nowhere*, 49.

Phillip Goff describes the combination problem in his article, “Experiences Don’t Sum,” where he states that for panpsychism to be valid, it must explain how “thousands of experience-involving ultimates come together in my brain to constitute the ‘big’ experience-involving thing that is the subject of my experience.”¹² Put another way, it is not apparent how microexperiences combine to yield macroexperience, for phenomenal consciousness does not *feel* like a thousand microexperiences—it feels like a unitary and indivisible awareness. Panpsychism must account for this in order to be considered viable. Otherwise, it is not useful as a theory of mind.

To make matters worse for panpsychism, the combination problem is not simply a problem. Instead, it is a cluster of problems. David Chalmers has done the work of teasing these out in his article entitled, *The Combination Problem of Panpsychism*, where he has identified three distinct sub-problems of the combination problem: the *subject combination problem*, which asks how microsubjects can combine to yield macrosubjects, the *quality combination problem*, which asks how microqualities can combine to yield macroqualities, and the *structure combination problem*, which asks how “microexperiential structure (and microphysical structure) combine to yield macroexperiential structure.”¹³ These sub-problems are distinct, but all broadly deal with how microexperiences combine to yield macroexperience. In the interest of avoiding compatibility issues, Chalmers recommends that any proposed solution to the combination problem address all three subproblems or at least specify which subproblem is being addressed.¹⁴

Below is an attempt at turning the *broad combination problem* into an argument, what I am calling *C*:

- (1) If panpsychism is true, then the combination of microexperiential ultimates must yield a discrete macroexperience.
- (2) It is never the case that the combination of microexperiential ultimates will yield a discrete macroexperience.
- (3) Panpsychism is false.

Implicit in premise 2 of *C* is the assumption that phenomenal consciousness is necessarily unified and that this is evident from how consciousness *feels*. This assumption is what I will object to in the coming section, as I hope to demonstrate that consciousness is not necessarily unified in the way that it feels. This is not to say that consciousness is disunified but that our perception of it is not infallible. Chalmers has suggested taking a deflationary account of the subject to tackle the subject combination problem (specifically, a sub-problem of the subject combination problem known as the *subject summing problem*), which is an unavoidable side effect of the coming argumentation. However, deflating the subjects is not my goal, as I hope to address all three sub-problems (in the form of the broad combination problem, *C*).

¹² Goff, “Experiences Don’t Sum,” 53.

¹³ Chalmers, “The Combination Problem for Panpsychism,” 5.

¹⁴ *Ibid.*, 6.

V. EXPERIENTIAL SUMMING

Premise 2 of *C*, which is based on Phillip Goff's broad assumption that experiences don't sum, is careful to distinguish microexperiences from macroexperiences. This is to say that the point of premise 2 is not to demand evidence of microexperiential summing from panpsychists but rather to question the emergence of discrete experiential agents at all, and regardless of size.

For this reason, evidence of any experiential summing—whether microscopic or macroscopic—could lead to a falsification of premise 2 and a subsequent invalidation of *C*. Clarification here is important because evidence of microexperiential summing is nowhere to be found. Macroexperiential summing, on the other hand, is something that has been demonstrated (although somewhat controversially) in psychological literature, namely, in the work of R. W. Sperry on so-called "split-brain" patients.¹⁵

Sperry conducted his research on the eleven patients of Phillip J. Vogel, who had received the complete surgical division of their corpus callosum in order to alleviate epileptic seizures.¹⁶ This procedure, called a *corpus callosotomy*, was the most radical brain bisection attempted in history at the time of publication. Therapeutic results of the surgeries were primarily positive, and of Vogel's eleven patients, only two still suffered from seizures after the procedure, the severity of which was seriously diminished.¹⁷ These results are fascinating in their own right, but the more relevant observations come from Sperry's experiments that tested the patients' unity of consciousness.

It should be noted that in a right-handed person, the left cerebral hemisphere is dominant and is the center of the person's cortical mechanisms for speech and writing; the opposite is true for a left-handed person. When the corpus callosum is severed, the two cerebral hemispheres cannot communicate through normal channels, such that when sensory inputs are segregated, the two hemispheres behave independently.

Sperry used various experimental methods, but the following is a notable example that demonstrates the peculiarity of the patients' behavior: a visual input was provided to the subject's left hemisphere via the right visual field, and he or she was able to describe the input verbally. However, when the same visual input was provided to the right visual field, the subjects asserted that they had not seen anything. Interestingly, if instead of asking the subject to verbally describe the object he or she has seen in his left visual field, they were asked to point to a matching image with their left hand, they would consistently point to the correct image (the same image they *say* they did not see).¹⁸ Sperry said this of his results:

“One of the more general and also more interesting and striking features of this syndrome may be summarized as an apparent doubling in most of the realms of conscious awareness. Instead of the normally unified

¹⁵ Roger Sperry's student, Michael Gazzaniga, has also produced work on split-brain patients that has been quite influential and enlightening. However, for brevity and simplicity's sake, it is not mentioned here. See bibliography.

¹⁶ R. W. Sperry, "Hemisphere Deconnection and Unity in Conscious Awareness," 723-733.

¹⁷ *Ibid.*, 723.

¹⁸ Sperry, "Hemisphere Deconnection and Unity in Conscious Awareness," 725.

single stream of consciousness, these patients behave in many ways as if they have two independent streams of awareness, one in each hemisphere, each of which is cut off from and out of contact with the mental experiences of the other.”¹⁹

Although these split-brain subjects behave in a way that suggests two consciousnesses, this is only the case under experimental conditions. Instead, subjects appear to have a normal unitary experience in their everyday lives.²⁰ Sperry interpreted this observation by concluding that, under normal circumstances, the two consciousnesses residing in the head of a split-brain patient run simultaneously to create one unified awareness.²¹

A commonly raised objection to this interpretation, which Sperry addresses in his article, is that the dominant hemisphere is the only conscious agent in the brain, while the minor hemisphere contains only functional properties. There is little reason to believe this, however, for although the minor hemisphere cannot express sensory inputs, it is still able to comprehend them.²² This is evidenced by the fact that when objects were flashed via projector to the left visual field and the subject was asked to pick out the object amongst others on a table that was hidden from view, the subject was able to locate it using his tactile senses, picking up and putting down objects until the correct one was found. When asked what he or she was holding, the subject consistently did not know how to respond. If the minor hemisphere is, in fact, an automaton without any stream of consciousness, then this behavior is not so easily explained.

In *Brain Bisection and the Unity of Consciousness*, Thomas Nagel expresses sympathy with two-minded hypotheses about split-brain research but does not believe that any whole number of minds can be discerned.²³ Nagel reasons that human assumptions about consciousness stand in the way of determining any true number of minds. After all, being a person certainly *feels* like being a single, unified stream of consciousness. If some evidence to the contrary could be found through introspection, would we even recognize it as such? Regardless of whether we can determine the true number of minds held by split-brain patients, we can certainly rule out the hypothesis that there is only one, as there appears to be no reason to believe this other than the fact that it has troubling implications. Even now, some 50 years later, the debate over the correct interpretation of split-brain research continues, but again the evidence for more-than-one-minded hypotheses seem far more compelling.

What does all this mean for panpsychism and the combination problem? If more-than-one-minded hypotheses are correct, then this is perhaps evidence of macroexperiential summing.²⁴ The argument below, what I am calling *S*, relates this evidence of macroexperiential summing to *C*.

¹⁹ *Ibid.*, 724.

²⁰ *Ibid.*, 728.

²¹ *Ibid.*, 732.

²² Sperry, “Hemisphere Deconnection and Unity in Conscious Awareness,” 731.

²³ Thomas Nagel, “Brain Bisection and the Unity of Consciousness,” 409.

²⁴ It has been suggested to me that more-than-one-minded hypotheses of split-brain research does not necessarily suggest macroexperiential summing, and may instead suggest that the global consciousness associated with the brain is an emergent consciousnesses. This view is expanded on in later sections of this article.

- (1) For panpsychism to be true, experiences must be able to combine to yield discrete experiences.
- (2) As evidenced by split-brain research, experiences can combine to yield discrete experiences.
- (3) Panpsychism can be true

There are four primary objections to this argument. Objection 1 is that experiential summing must be displayed in normal brains in order to criticize premise 2 of *C* legitimately, while *S* only accounts for abnormal brains. Objection 2 is that *S* is too different from *C* to be useful (*C* addresses microexperiences while *S* addresses experience broadly). Objection 3 is that while *S* displays experiential summing, it still does not create a positive account of *how* experiences might sum. Finally, objection 4 is that *S* is too deflationary to the normal conception of consciousness. Let us address these objections sequentially to see how *S* holds up.

VI. OBJECTIONS AND REPLIES

Objection 1 claims that experiential summing in abnormal brains tells us nothing about the possibility of experiential summing in normal brains. This objection assumes that panpsychism, if true, is a fundamental fact about reality, and so evidence in its favor cannot be forced by surgical alterations to the brain. Instead, evidence must come about more naturally.

Objection 1 is correct to assume that panpsychism, if true, must be a fundamental fact about reality but misses the point of premise 2 of *C*. Premise 2 of *C*'s only claim is that experiences never sum, and so this premise can be invalidated by any evidence of experiential summing, regardless of the medium. This point should be enough to invalidate objection 1, but one could also reply by appealing to *multiple realizability*, which is the thesis that mental states ought to be attributed to any entity that properly exhibits them—not only those with brains. Chalmers paraphrases multiple realizability by saying that “states such as pain cannot be identical to any particular brain states, since a creature such as a Martian might have pain without having the brain state in question.”²⁵

Hilary Putnam famously appealed to multiple realization in his 1973 article, *The Nature of Mental States*, in which he objected to *identity theory* (a materialist theory that states that mental states are identical to brain states).²⁶ Putnam appealed to multiple realizability as a means to defend *functionalism* (the theory that mental states correspond to functional states), but one need not be a functionalist in order to appeal to multiple realizability. On the contrary, multiple realizability is compatible with panpsychism and is one of its strengths.²⁷ If one is committed to multiple realizability, then normal brains should have no significance.

²⁵ Chalmers, *Philosophy of Mind: Contemporary Readings*, 4.

²⁶ Hilary Putnam, “The Nature of Mental States,” *In Philosophy of Mind: Contemporary Readings*, 73-79.

²⁷ This point has been objected to on the grounds that, according to panpsychism, a brain made of anti-matter would lack consciousness, and thus panpsychism is not compatible with multiple realizability. I suppose that this depends on how consciousness would be identified in all possible panpsychist worlds; if panpsychism is true in this world, would there not be some analogous anti-experience that would hold in an anti-matter world?

Although objection 1 seems to fail for the reasons given above, it does ask a good question: How can abnormal brains tell us anything about normal brains? Perhaps the answer is that the disparity between the two is not as significant as it seems. For instance, in the case of a normal brain, the corpus callosum serves as a communication device between the two hemispheres. In the case of a split-brain under non-experimental conditions, this communication is carried out by overlap of sensory fields. In both cases, there appears to be a person with a unitary consciousness. It seems then that under normal circumstances, there is little distinction between the two and that the two hemispheres of a normal brain contain their own discrete consciousnesses in a way that is similar to a split-brain.

One may object here that if the mind associated with a normal brain is the sum of two consciousness, why then has this not been revealed through introspection? This objection assumes that introspection can reveal *everything* about the nature of consciousness, an assumption that is most certainly false. For example, introspection does not reveal the fact that brains have anything to do with cognition at all. The cognitive scientist, Douglas Hofstadter has argued along similar lines, saying that “there could be some high-level way of viewing the mind/brain, involving concepts which do not appear on lower levels, and that this level might have explanatory power that does not exist—not even in principle—on lower levels.”²⁸ Perhaps one who makes this objection also assumes that a mental event arising in one hemisphere would seem different from a mental event arising in the other. This is possible, but the two hemispheres also go everywhere together, meet the same people, and do nearly everything in tandem. Therefore, if introspection *could* reveal everything about consciousness, it does not seem likely that these mental events would differ significantly enough for one to notice.

Objection 2 was touched on in section V but will receive more attention here; namely, that *S* only addresses macroexperiential summing, and that microexperiential summing must be demonstrated to invalidate premise 2 of *C*. One might say that macroexperience is so different from microexperience that their properties cannot be adequately compared. For example, the combination of two cerebral hemispheres might reveal nothing about the combination of quarks or atoms. This seems at least partially true—whatever experience a quark may have is undoubtedly dissimilar from that of a human, but this objection claims that experience *itself* can vary between conscious agents; how is this possible? The mental process of experience, or what-it-is-likeness, cannot possibly vary; otherwise, we would be talking about something else. What it is like to be a quark may be different from what it is like to be a person, but only by nature of the contents or qualities of experience, not experience itself.

This is to say that objection 2 is actually a reformulation of the quality combination problem. *S* has little to say about how qualities might combine, but it does assume that it is possible. For instance, take the hypothetical case of a split-brain patient who is given a visual stimulus to his dominant hemisphere, and some other visual stimulus is given to his minor hemisphere. In this case, both hemispheres are experiencing different qualities of consciousness, but the nature of experience, *viz.*, *awareness*, is unchanged. In this case, how do the

²⁸ Hofstadter, D. (1979) *Gödel, Escher, Bach: An Eternal Golden Braid*, 409.

qualities combine to yield a unitary experience? Under experimental conditions, no experiential combination occurs, but under normal conditions, experiential combination occurs via sensory overlap that facilitates communication between cerebral hemispheres. In a normal brain, this same combination takes place via communication through the corpus callosum.

Again, *S* says nothing about *how* experiences might combine but asserts that they must somehow.

Objection 3 states that even though *S* demonstrates evidence favoring experiential summing, it still does not create a positive account for how experiential summing might occur. This is a fair objection; so far, this article and *S* have only discussed the possibility of two minds associated with cerebral hemispheres that combine to yield the global consciousness associated with the brain. However, a positive account of panpsychism would have to discuss the possibility of an enormous number of minds associated with the brain.

A complete panpsychist account of the brain will not be attempted here, but if one takes the implications of split-brain research seriously, then the framework of such an account begins to take form. Nagel once noted that if one accepts two-minded hypotheses of split-brain research, “why then should we not regard *each* hemisphere as inhabited by several cooperating minds with specialized capabilities? Where is one to stop? If the decision on the number of minds associated with a brain is largely arbitrary, the original point of the question has disappeared.”²⁹ In other words, why should we stop at *two* when counting consciousnesses and not attribute discrete consciousnesses to all subsystems in the brain? For Nagel, this is enough reason to reject two minded-hypotheses, but for a defender of panpsychism, it should be a reason to accept them. Let us now take Nagel’s point seriously and see how it may lead to a positive account of experiential summing in the brain.

We can begin by saying that the global consciousness associated with the entire brain is a combination of *overlapping* consciousnesses associated with subsystems, sub-subsystems, all the way down to fundamental constituents of matter. Any microexperiential unit that resides in this system, such as an atom or quark, would be an independent experience *and* be associated with multiple subsystems. This may seem like a problem—how can discrete experiences be held by multiple experiential units? Recall, however, that this same phenomenon occurs in split-brains with no difficulty; the left hemisphere is a discrete consciousness and is *also* part of the global consciousness associated with the brain. If a hemisphere can behave in this way, viz., with mental events belonging to multiple experiential units, then it follows that other subsystems can behave similarly.³⁰ Additionally, it is not clear what aspects of experience will be accessed by the ‘top level’ at any given time. If, for example, some portion of subsystems experience red and only red while other subsystems experience green and only green, then perhaps ‘the top level’ is unable to determine which microexperiences are to be available to itself.

²⁹ Nagel, “Brain Bisection and Unity of Consciousness,” 413.

³⁰ Russellian monism does not choose between the possibility of panpsychism and 31 panprotopsychism, so if panprotopsychism is correct, then it is not necessary that all subsystems in the brain be conscious. This may be helpful to those who do not feel comfortable ascribing consciousness to very primitive subsystems, like the brainstem.

The result of this view is a conception of the brain that includes an enormous number of minds, associated with fundamental constituents of matter, that combine to yield even more minds at every “level” of combination, and culminating at the level of the entire brain. Based on the previous discussions in this article, this seems perfectly conceivable and perhaps even mappable with the help of computers, but is the entire brain the only possible culmination of the conscious system? This is a difficult question to answer and one that probably deserves the courtesy of its own article, but the answer is not obviously “yes.” After all, are we to include the brainstem in this conscious system? The spinal cord? Spinal nerves? If consciousness or proto-consciousness is associated with all matter, then where is it to end, and why should the brain be given special status? At any rate, the account put forward in this section is, of course, incomplete and only a suggestion to someone who may wish to create a complete panpsychist account of mind. Ultimately, objection 3 does not have a conclusive reply, and so without a complete positive account of panpsychism, it remains an open question as to whether *S* is viable.

Objection 4 claims that *S* is too optimistic about deflating the normal conception of consciousness. This objection is especially appropriate bearing in mind the suggestions found in the previous section, and the way that consciousness *feels* should undoubtedly be considered when discussing it. However, it is possible that this is not all that should be considered, so any comprehensive theory of mind must be somewhat deflationary.

It is hard to square a deflationary view of consciousness, self, or subjectivity with our normal experience, but many philosophers have defended a deflationary view for various reasons. Daniel Dennett is one such philosopher, and in his article, *The Self as a Center of Narrative Gravity*, he argued that the self is not the unmoving center of consciousness that it seems, but a fictional character, dreamt up by whatever processes are going on in the brain. Dennett writes that we

“find ourselves engaged in all sorts of behavior, more or less unified, but sometimes disunified, and we always put the best “faces” on it we can. We try to make all of our material cohere into a single good story. And that story is our autobiography[...] we are not the captains of our ships; there is no conscious self that is unproblematically in command of the mind’s resources. Rather, we are somewhat disunified. Our component modules have to act in opportunistic but amazingly resourceful ways to produce a modicum of behavioral unity, which is then enhanced by an illusion of greater unity.”³¹ Although Dennett here refers to the unity of *self* (rather than the unity of consciousness),

the two sorts of unity seem closely linked, for there must be a self, or some other organizing feature, that stands in relation to experience. At any rate, Dennett’s view is that disunity need not be demonstrated only in exceptional cases like split-brains but that it is apparent in all sorts of normal behavior. Everyone can recall times when they have behaved ‘out of character’; dispositions change throughout people’s lives, varying from day to day, hour to hour, and minute to minute. People

³¹ Daniel C Dennett, “The Self as a Center of Narrative Gravity,” 103-115.

behave in disunified ways *all the time*, so why should this not be taken into account?

Dennett's assertion that the self is disunified seems correct, but then what is one to make of the sensation of unity? Perhaps maintaining this sensation of unity while acknowledging that it is not *necessarily* veridical is enough to satisfy our desires to resist deflationary accounts of consciousness while squaring it with *S* and the panpsychist account of consciousness in the previous section. In other words, even if the mind is somewhat disunified, why should there not be some feature of it that plays the role of projecting the illusion of unity, as Dennett suggests?

VII. KANT AND MENTAL UNITY

Thus far, the topic of mental unity has been mentioned only in passing, but clearly, unity is deeply coupled to the concept of combination. It is not altogether evident how the two concepts are linked, but let us conjecture that however experiences are combined, the resulting macroexperience must be unified to be considered valid. Let us then examine unity more directly.

Is the view of unity presented in the previous section powerful enough to square the normal conception of consciousness with *S*? If one agrees with David Hume that the mind is a "bundle or collection of different perceptions," then there is no need to probe further.³² Problems begin to arise, however, when *S* is confronted with philosophies of mind that see unity as necessary, such as the philosophy of Immanuel Kant. Kant's conception of unity—what he calls the *transcendental unity of apperception*—is roughly that the mind must be unified because cognition is otherwise not possible.³³ This is clearly not *prima facie* compatible with the conception of unity that *S* implies, but is there any way for defenders of panpsychism to resist Hume's position that the mind is necessarily disunified?

First, we should ask how one is to think of mental unity. When one examines her entire field of consciousness, it may contain multiple objects of various colors and textures existing in a single space, and there may be sounds like that of birds chirping or cars passing by. Here, we can say that there is unity of *objects* of experience. Despite having different qualities (color, texture, shape), these objects are unified and experienced as singular, and they are presented in a unified field of consciousness. This sort of unity is important for our purposes, i.e., the unity of objects of experience. Still objects of awareness must presumably be presented to some sort of subject—how then is this subject unified?

Tim Bayne and David Chalmers write in *What is the Unity of Consciousness?* that subject unity occurs "if a set of conscious states are had by a subject at a time."³⁴ For example, the scene described in the previous paragraph possesses subject unity because a single subject holds the experiences. Although Kant discusses several kinds of unity—both of objects and the subject who experiences them—it is subject unity that he sees as necessary for cognition. He writes that "there can take place in

³² David Hume, *Treatise on Human Nature*, ed. Jonathan Bennett (Jonathan Bennett, 2017), 131.

³³ Immanuel Kant, *Critique of Pure Reason*, trans. Werner S. Pluhar (Indianapolis, Indiana: 34 Hackett Publishing Company, 1996). A107.

³⁴ Tim Bayne and David Chalmers, "What is the Unity of Consciousness?," 5.

us no cognitions, and no connection and unity of cognitions among one another, without that unity of consciousness which precedes all data of intuitions, and by reference to which all presentation of objects is alone possible.”³⁵ This passage clarifies that although Kant thinks that objects of experience must have unity, they must be presented to a single, unified subject to be experienced at all.

S claims that experiences can combine to yield discrete experience and that this can be observed in split-brain patients who hold some unknown number of subjects while being a single, unified subject. Is there some way to square this with Kant’s position? One way is perhaps to consider a sort of unity called *subsumptive unity* that is suggested by Bayne and Chalmers.³⁶ Subsumptive unity refers to a subject’s total field of consciousness; it is not the conjunction of all objects of experience but a discrete conscious state that subsumes them.

It seems plausible that this sort of unity exists, but it does not apply to the subject. Instead, it assumes that the entire consciousness field is necessarily projected to a united subject. However, if one is willing to accept the emergence of a subsumptively unified conscious field, then perhaps the principle can also be applied to subject unity. The previous section hypothesized that something like subsumptive unity occurs at every level of experiential combination in the brain—from fundamental constituents of matter to the brain as a whole. Perhaps the bundle of perceptions that Hume identified with the mind is subsumed to become a single subject rather than a conjunction of perceptions. If this is the case, then *subsumptive subject unity* could occur in split-brains functioning under non-experimental conditions to subsume the two hemispheres.

It is doubtful whether Kant would find this solution satisfactory, as he was known for his absolutism. Still, it may be attractive to others, especially defenders of panpsychism, who may also wish to keep the mind unified. The claim in this section is different from the one found in the previous section, namely, that the mind is not necessarily unified but projects some appearance of unity. It is not clear which claim, if any, is correct.

VIII. CONCLUSION

The argumentation in this article has led to two possible outcomes—neither of which are free from problems of their own. On the one hand, I have asserted that the mind may possess subsumptive subject unity, which some may argue appeals to brute emergence. On the other hand, I have claimed that the mind may not be necessarily unified but at least appears to be; a claim that is difficult to square with cognitive theories like Kant’s (and could lead to unwanted moral implications). A closer examination of both these claims is needed.

Despite the uncertainty surrounding these implications, this article should hopefully have achieved its goal of injecting skepticism into the combination problem. The assumption that experiences do not sum appeals to our commonplace notion of mental unity. However, when confronted with the example of split-brains, which appear to demonstrate experiential summing—the assumption begins to lose its force.

³⁵ Immanuel Kant, *Critique of Pure Reason*, trans. Werner S. Pluhar (Indianapolis, Indiana: 34 Hackett Publishing Company, 1996). A107.

³⁶ Bayne and Chalmers, “What is the Unity of Consciousness?” 5.

BIBLIOGRAPHY

Bayne, Tim and David Chalmers. "What is the Unity of Consciousness?." Essay. In *The Unity of Consciousness: Binding, Integration, Dissociation*, edited by Axel Cleermans. Oxford: Oxford University Press, 2003.

Chalmers, David. "Foundations." Introduction. In *Philosophy of Mind: Contemporary Readings*, edited by David Chalmers, 1-9. New York, New York: Oxford University Press, 2002.

———. "Panpsychism and Panprotopsychism." Accessed February 15, 2021. <http://www.consc.net/papers/panpsychism.pdf>.

———. "The Combination Problem for Panpsychism." Accessed February 15, 2021. <http://www.consc.net/papers/combination.pdf>.

———. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press, 1966.

Dennett, Daniel C. "The Self as a Center of Narrative Gravity." Essay. In *Self and Consciousness: Multiple Perspectives*, edited by Frank S Kessell, Pamela M Cole, and Dale L Johnson, 103-15. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1992.

Gazzaniga, Michael. *The Social Brain*. New York, NY: Basic Books, 1985. Goff, Phillip. "Experiences Don't Sum," *Journal of Consciousness Studies* 13, no. 10-11 (2006).

Hofstadter, Douglas R. Gödel, Escher, *Bach: An Eternal Golden Braid*. New York: Basic Books, 1999.

Hume, David. *Treatise on Human Nature*. Edited by Jonathan Bennett. Jonathan Bennett, 2017. Jackson, Frank. "Epiphenomenal Qualia," *Philosophical Quarterly* 32, no. 127 (1982).

Kant, Immanuel. *Critique of Pure Reason*. Translated by Werner S Pluhar. Indianapolis, Indiana: Hackett Publishing Company, 1996.

Nagel, Thomas. "Brain Bisection and the Unity of Consciousness." *Synthese* 22, no. 3-4 (1971): 396-413. <https://doi.org/10.1007/bf00413435>

Nagel, Thomas "What is it Like to Be a Bat?," *The Philosophical Review* 83, no. 4 (1974).

Nagel, Thomas. *The View from Nowhere*. New York: Oxford University Press, 1989.

Putnam, Hilary. "The Nature of Mental States." Essay. In *Philosophy of Mind: Contemporary Readings*, edited by David Chalmers, 73-79. New York, New York: Oxford University Press, 2002.

Sperry, R. W. "Hemisphere Deconnection and Unity in Conscious Awareness.," *American Psychologist* 23, no. 10 (1968): 723-733. <https://doi.org/10.1037/h0026839>.

Plato and the Problem of Tragedy

Weiming Sheng
Vassar College

FOREWORD

Editor Sophia Gottfried

Why is tragedy so popular when the genre itself is focused on engendering pain? Why, when there is such emphasis on avoiding pain in day-to-day life, do humans flock to tear-producing images of suffering?

The paper below, "Plato and the Problem of Tragedy," shows how in Book X of *The Republic*, Plato tackles the question of why we desire art that induces pity and the issues that such an inquiry poses for his tripartite theory of the soul. Plato's solution, outlined in the piece, is proposing an amendment to his traditional tripartite view of the soul, adding another part just for pity.

We selected this paper because of the creative and rigorous reading of *The Republic*. The paper succeeds in its bipartite purpose: presenting a new and interesting interpretation of Book X in relation to the view of the soul of Book IV, and also setting Plato's view in the historical and contemporary discussions of tragedy. Sheng spotlights the oft-neglected anti-rational pitying portion of Plato's soul, akin to the outcasts and unwanted of Kallipolis, showing us the deficiencies in the Platonic conception of psychological justice. The paper ends with highlighting issues regarding Plato's view of tragedy, and gives us considerations to further question why we seek, in defiance of rationality, that which makes us weep.

ABSTRACT

Plato's critique of poetry in Book X of the *Republic* is scathing but insightful. In this essay, I understand his critique as a response to the problem of tragedy, which asks how an authentic experience of distress, for the audience of a tragedy, turns into a genuine pleasure. I first compare Plato's answer with Jean-Baptiste Du Bos' and David Hume's. I then argue that the implication to Plato's answer is a revision of the moral psychology presented in Book IV. Finally, I propose a motive for Plato's critique of tragedy by examining the inadequacy of the psychic conception of justice.

I. THE PROBLEM OF TRAGEDY

1965 caught Bob Dylan particularly feisty with the press. In a remarkable interview with Horace Judson from the *Time*, Dylan asked the reporter, "Do you think somebody would go see somebody if they didn't want entertainment?" Judson said no. Dylan went on, "Who, now who, wants to go get whipped, you know, and if you do wanna go get whipped, hey, aren't you really being entertained?" To find entertainment from getting whipped sounds a bit weird, but the phenomenon Dylan was referring to is ubiquitous. Many consider Stanley Kubrick's *The Shining* to be one of the best films ever made, and the horror/thriller genre has enjoyed popularity for decades; every fun amusement park has a rollercoaster, or something similar to it, which renders a significant level of physical unease to the passengers; there are people in all parts of the world who love spicy food, but the capsaicin in them causes real somatic pain from the moment they enter the mouth until they leave the body. In all these cases, people have a good time quite inexplicably from something painful and almost torturous.

This applies also to the pleasure one derives from watching or reading tragedies.¹ David Hume eloquently describes the phenomenon in the following:

‘It seems an unaccountable pleasure which the spectators of a well-written tragedy receive from sorrow, terror, anxiety, and other passions that are in themselves disagreeable and uneasy... They are pleased in proportion as they are afflicted, and never are so happy as when they employ tears, sobs, and cries, to give vent to their sorrow, and relieve their heart, swoln with the tenderest sympathy and compassion.’ (221).

Similarly, August W. Schlegel asks, “Why does Tragedy select subjects so awfully repugnant to the wishes and the wants of our sensuous nature?” (67). The core of this problem of tragedy is how an authentic experience of distress, for the audience of a tragedy, turns into a genuine pleasure.

As Schlegel goes on to say, this is a question that “has often been asked, and seldom satisfactorily answered” (67). Indeed, the problem of tragedy features quite prominently in contemporary aesthetics debates on the topic, but the origin of the problem is often traced back only to Aristotle’s *Poetics*; if Plato is mentioned, it is usually only in connection to that work.² This is very unfortunate, for not only was Plato perhaps the first to have posed this question, but his answer to it remains, I think, among the most interesting and provocative. In this essay, I attempt to make clear the substance, motive, and implications of Plato’s answer to this question in Book X of the *Republic*. In §2-3, I give an overview of how Plato’s answer, on first inspection, differs from some other responses in accepting that we enjoy tragedies for their subject matters and in being evaluatively critical to the project of tragic poetry. In §4-7, I examine what Plato’s answer implies for his theory of the soul. If my arguments are right, in Book X Plato posits a fourth part of the soul, an anti-rational pitying part, in addition to the three parts already introduced in Book IV. Then I propose a potential motive behind Plato’s critique in §8-10, which has to do with a shortcoming in Plato’s psychic conception of justice. This motive which I have speculated, as I will explain and conclude in §11, is useful to the extent that it repudiates those who do not take Book X’s arguments very seriously by treating it as a mere appendix or an exercise of irony.

II. THE DISTRESS OF TRAGEDY

If some of what is depicted in tragedies were to happen in our own life, we would certainly not like it. It thus appears that the reason why we welcome tragedies cannot primarily be because of its subject matter (which causes us distress), but has rather something to do with its poetic/dramatic form. This line of reasoning is pursued by the French aesthetician Jean-Baptiste Du Bos, who writes that “The grief [of tragedies] that appears is only, as it were, on the surface of our heart, and we are sensible, that our tears will finish with the representation of the ingenious fiction that gave them birth” (25). The superficial and temporary distress of tragedies counts towards pleasure because it rids us of languor and boredom, compared to which it is better by far.

¹ The concept of “tragedy” or “the tragic” is notoriously difficult to pin down. Luckily for me, when speaking of tragedies, Plato seems to have in mind, very specifically, the works of Homer and his contemporary tragedians (605c).

² See, e.g., Neill 366-367.

Hume, while rejecting Du Bos' explanation, agrees that we enjoy tragedies for their poetry. He speaks of a *conversion* of negative emotions: "The impulse or vehemence arising from sorrow, compassion, indignation, receives a new direction from the sentiments of beauty" (225). That is, we find genuine enjoyment in perceiving "the exercise of noble talents," and this, being "the predominant emotion" in our mind, transforms the negative emotions for the subject matter of tragedies into pleasure, while absorbing their intensity.

These two responses to the problem of tragedy serve as useful contrasts to Plato's position. In effect, Plato rejects the premise which begins their reasoning: namely, that we do not welcome tragedies because of the distress associated with their subject matter. This is not to say that he does not recognize the felicity and power of poetry: at 601a-b, Socrates³ claims that people believe poets to have knowledge just because they "speak with meter, rhythm, and harmony". "So great," Socrates warns, "is the spell of these things by nature."⁴ And yet the "chief charge" Plato prepares against the poet is concerned with their subject matter, and it is through this, not their poetic form, that tragedies "corrupt even decent people" (605c).

There are two objections to the claim that we welcome tragedies also for the distress caused by their subject matter. The first is that, since we certainly do not prize distress when experienced outside the context of tragedy, but we in fact detest it, why should we value tragedy which reproduces these disturbances? For example, we would certainly be quite disturbed if we heard that a son killed his father and married his mother, and we would not think that this disturbance is a good thing—why then should *Oedipus* receive accolades for reproducing the same effect in us? This question is dodged by Du Bos and Hume because they claim that the subject matter of tragedies arouses considerably different reactions than its counterpart in real life.

Plato, however, bites the bullet and does so readily. He believes that we should not treat the distress caused by tragedy any differently than distresses in general, and since in general distresses are to be suppressed and "we pride ourselves if we are able to keep quiet and master our grief" (605d), it makes no sense that we instead let ourselves go in theaters and succumb to lamentations. Thus, Plato can respond to this objection by simply accepting its consequence, namely, that we should *not* value tragedies for the distress that their subject matter evokes. If this means we should not value tragedy, so be it (608a-b).

The second objection is more difficult for Plato to deal with. If tragedies cause us genuine distress, then why *do* we find genuine pleasure from the experience? This fact is acknowledged by Plato: Socrates points out that when a bard imitates a grieving hero, "we *enjoy* it and give ourselves over to it" (605d). How can this phenomenon be explained if poetry causes us the same sort of distress as events in real life? This objection is at the core of the problem of tragedy, and Plato makes it more difficult for himself by rejecting Du Bos and Hume's premise; however, I think the main interest to his critique of poetry comes from the fact that he

³ Socrates is the main character in most of Plato's dialogues. To what extent, and in which dialogues, do his views represent Plato's are under dispute. I assume in this paper that in the *Republic* he is used as a mouthpiece for Plato's own views, and thus I sometimes use "Socrates" and "Plato" interchangeably.

⁴ Translations from C. D. C. Reeve with some variations.

decided to deal with this difficulty head-on.

The first objection is evaluative, and Plato could resolve it quite simply by agreeing with the evaluative consequence. The second objection is descriptive, and Plato cannot resolve it by claiming a counterfactual.⁵ Instead, he must give an account of the nature of these pleasures that even “decent people” get from tragic performances. To do this, Plato offers a very suggestive psychological account of the experience of being affected by tragedies. To see whether Plato was able to account for the tragic pleasures, it is necessary to go over this psychological aspect of his critique of poetry.

III. OUTLINE OF PLATO’S RESPONSE

In summary, the following is how Plato explains the pleasure we experience from tragedies in Book X of the *Republic*: This pleasure belongs to an inferior part of ourselves which derives satisfaction from wailing and lamenting, a part usually kept in check, since social norms and reason require for us not to succumb completely to grief even when we suffer misfortunes; however, when we are watching a tragedy, since it is concerned with someone else—often one with a claim to goodness—we let our guard slip and think that it is not shameful to give in to laments. The pleasure of tragedy, therefore, comes from the release of this usually-restrained pitying part.

This is surely an interesting idea, but it is not enough. To explain the pleasure we feel for tragic distress by positing a part of our soul that just happens to house such a desire is plainly question-begging. For Plato’s explanation to be more substantive, he needs to offer a more general theory or a model, in the context of which this part that takes pleasure in tragedies can be situated and explained. This, I believe, he in fact accomplishes in Book X. The theory he offers amounts to a *quadripartition* of the soul, which is a significant revision of the psychological theory presented in Book IV, where the soul is divided into three parts or subjects.⁶

The theory of the soul in Book X begins with two characterizations of the fourth part which is affected by tragedy. I now move on to an overview of some previous attempts in the secondary literature at identifying this part with one of the other three parts of the soul given in Book IV: reason, spirit, and appetite.

IV. THE INFERIOR PART

After establishing that “the imitator knows nothing worth mentioning about the things he imitates” (602b),⁷ Socrates goes on to ask Glaucon, “on which of the elements in a human being does imitation have its effect?” (602c). A few lines

⁵ Another way to put this is that the first objection is moral-aesthetic, since it is concerned with the *values* of tragedy, in itself and in the context of arts; the latter objection is psychological.

⁶ For what a division of the soul more specifically means and entails, see Bobonich 219-235.

⁷ Involved here is the metaphysical critique of poetry (595c-602b). The most important point is that poetry is claimed to be thrice removed from truth, because it represents an *appearance* of the craftsmen’s products, which in turn are only representations of the Form (599a). This thesis is significantly more complicated than it sounds: Alexander Nehamas has pointed out that the product of the mimetic artists and the thing imitated are ontologically equivalent (69). Jessica Moss goes further to recognize both the ontological and the metaphysical distinction between what is imitated (i.e. an appearance) and the product of the craftsmen (*What*, 420-422). See their articles for further discussions.

down he answers this question himself: “imitation really consorts with an element in us that is far from wisdom” (603b).⁸ Later on, at 604d, Socrates decries this part further as “irrational”, “lazy”, and “cowardly”. Previously, in Book IV, Socrates divides the soul⁹ into reason, spirit, and appetite.¹⁰ Though at times he seems to suggest that this division does not have to be exhaustive,¹¹ there is no reason *prima facie* to think that he is speaking of a part here which deviates from that division.

Indeed, the commentators have attempted in various ways to identify this “lower part of the soul”. James Adam says that it “appears sometimes as the ἐπιθυμητικόν [appetitive part], sometimes as a degenerate form of the θυμοειδές [spirited part]” (406).¹² Nehamas, on the other hand, identifies it as “belonging to the rational part of the soul”; specifically, it is characterized by “the uncritical acceptance of the senses’ reports” (65).¹³ And if none of these identifications should appear quite right, one could claim instead that the accounts of the soul offered in Book IV and Book X are inconsistent. This is in effect how Annas deals with the psychology in Book X: “Plato in fact is very vague here about the soul’s lower part...In this passage [602c-603a] he always refers to it simply as the worthless part, keeping in the background the fact that to be consistent with its roles elsewhere it would have to be the *desiring* part” (340).¹⁴

Thus, it looks like every interpretative option has been tried, but none, as suggested in the notes and as I think will become clear, is quite right. Socrates offers two characterizations of this part of the soul, and I think it makes the most sense to take the first characterization as an illustration or heuristic analogy of the part, which is really described by the second characterization (602c-603b, 603b-605c).¹⁵ I will begin with the first characterization.

V. FIRST CHARACTERIZATION

In the first characterization, the part of the soul in question is characterized as the part which responds to optical illusions such as the *trompe l’oeil* paintings

⁸ As S. Halliwell and others have noted, the Greek here and in Book IV is deliberately ambiguous, for the “parts” or “elements” of the soul are spoken of mostly with indefinite pronouns (*Plato*, 134).

⁹ The ancient Greeks’ idea of the soul (ψυχή) is free of much of the conceptual baggage that it carries for us. Julia Annas suggests that “Plato’s talk of ‘soul’ is rather like our talk of someone’s ‘mind’ to refer to certain mental phenomena without committing ourselves thereby to any particular theory about the mind” (124). I accept this: hence, in this essay, when I speak of the “soul,” I mean, broadly, “that part of a man’s personal life with which he identifies *himself*” (Crombie 34).

¹⁰ The division is usually taken to be “a theory about the nature of the embodied human soul and a theory of human motivation” (Lorenz, *Plato* 243).

¹¹ For example, he allows that there might be “some others in between” (443d).

¹² For the latter, Adam cites 604e, 606a, and 606d, though it’s not obvious to me how any of these passages supports the claim. Moss too sees 606d as in favor of this thesis, but here manifestly appetite and spirit are discussed *as parallels* to the part in question and not as identical to it (*Appearances* 42-43).

¹³ Of course, this conception goes very badly with the characterization Socrates gives of it at 604d, cited already at the end of the last paragraph. While it is true that 602b-605a is concerned only “with the subject matter of the poem itself,” the fact that mimetic poetry has this subject matter is clearly intended by Socrates as evidence for his claim that it “arouses”, “nourishes”, and “strengthens” this lower part of the soul, and thereby “destroys the rational part” (67, *Plato* 605b).

¹⁵ Plato holds the stronger claim that these characterizations, in fact, describe *the same part* (605b-c), and not just one part and something else which is similar to it. However, the arguments he gives do not show identity but only similarity in several respects. Since he also holds the weaker claim that the two are similar (605a), I will reject, along with M. F. Burnyeat (224-226), the stronger claim as ungrounded.

(602c). The first thing to note about it is that, even after reason has calculated the correct proportion of the object, “the opposite appears to it at the same time” (602e). The important point is not that the part shares with reason a cognitive capacity, but that it doesn’t listen: despite reason’s best effort, it is not swayed. Moreover, manifestly, it is not because an individual is particularly dissolute or irrational that this lower part in them does not listen to reason: rather, surely the just person also experiences optical illusions, in the same way that an unjust person experiences it. Then, this part is different from spirit and appetite in this respect, for both spirit and appetite are tamable and do in fact listen to reason in the soul of the just person. It is also obviously different from reason, for it is impossible for reason not to listen to reason.

Moreover, in Plato’s ideal city, the productive and the auxiliary classes make not only positive contributions but are necessary components of it (433a-b). Analogously, for someone to be just, it is not enough to be philosophic, but some appetitive and spirited desires should also be present in them (443c-444a). The part of the soul which produces optical illusions, however, makes no contribution to the justice of the person, and it is difficult to see what place its analogous class would have in the Kallipolis.¹⁶ Thus, the lower part of the soul in the first characterization is different from the three parts offered in Book IV, for, unlike them, it is *placeless* in the ideal state.

The first characterization is thus entirely negative: we see something in our soul that 1) differs in its belief from reason’s calculation, 2) insists in its error against reason, and 3) is placeless in the ideal city. To the extent that reason calculates correctly, it stands in opposition to reason; and to the extent that spirit and appetite do not insist in their errors against reason in a well-adjusted person, it stands in opposition to them too. To the extent that the analogous classes of all three other elements are necessary components of the ideal state, this part stands in opposition to them all. Since all three characteristics involve opposition to reason in some way, I will refer to a part in our soul which has these aspects anti-reason.

Socrates claims that “the poet is like the painter in associating with a part of the soul that is similarly inferior rather than with the best one” (605a). If the first characterization is an illustration of the part in us which responds to tragedies, it is plausible to think that the illustration consists of the two parts sharing these same negative aspects, for these seem to be what its “inferiority” would consist in. Do these aspects then come out in Socrates’ second characterization?

VI. SECOND CHARACTERIZATION

The second characterization directly describes the part responsible for tragic pleasures by deeming it as the part which tragedies imitate and thereby strengthen (605a). It has “an insatiable hunger for weeping and lamenting—since that is what it has a natural appetite for” and is “the very factor that gets satisfaction and

¹⁶ In the ideally just person, there should be perfect harmony and a perfect “doing-one’s-own” for each part of the soul (443c-444a). Since the proper role for reason is to rule, this prerogative should mean that, whenever reason commands, say, an appetitive desire to desist, it desists. Appetite and spirit are overcome and preserved, so to speak. Only the lower part of Book X has been described as a human “πῶθημα”—rendered by most translators as “weakness”—whereas neither spirit nor appetite *tout court* has been described in this way (602c).

enjoyment from the poets” (606a). I will follow Plato in calling it “the pitying part” and sometimes simply “pity” (606b).

I think the pitying part is anti-reason since the three criteria from the first characterization apply to it. At 604a, Socrates claims that, for the decent man who has suffered a considerable loss, “reason and law tell him to resist, while what urges him to give in to the pains is that very weakness (πάθος).”¹⁷ While both reason and law are cited as what opposes the weakness, the content of the law given at 604b-c is completely rational. Plato thinks that, according to rational assessment, “it is best to keep as quiet as possible in misfortunes and not get irritated, since what is really good or bad in such things is not clear” (604b). Just as reason shows the correct proportion of a thing looked at from far away, here reason shows the correct amount of grief one should allow themselves, which is, admittedly, not very much at all. If this is so, then the pitying part meets 1), for it also differs in its belief from reason’s calculation, since it believes the misfortune to justify a greater amount of grief than it rationally does.

Now, after giving its assessment, reason desires to move on (604c). However, the pitying part of our soul desires precisely to dwell further on the pain: it “leads us to *recollections* of our suffering and to lamentations, and is insatiable for these things” (604d). This is true even for the decent man (ἐπιεικής), whose soul is, supposedly, in good order.¹⁸ It, therefore, meets 2).

On 3), if people analogous to the pitying part are present in Plato’s ideal city, they must belong to one of the three classes. Now, they obviously do not belong to the class of the rulers, nor do they belong to the class of auxiliaries, whose education is precisely to prepare for them to mourn as little as possible when suffering misfortunes (386a-388d). Since the third class is characterized primarily as “providers of upkeep and wages”, they probably do not belong here either. Indeed, there is reason to think that the poets (those among them to be exiled) are the very ones who would represent the pitying part in the political analogy,¹⁹ and if this is so, their banishment in Book III and Book X would mean that the people analogous to the pitying part are placeless in the ideal state. Furthermore, earlier in this section, I noted that Socrates describes the pitying part as a “weakness” (πάθος). Just as in the case of optical illusions, we would rather not have to suffer a weakness, so that in the ideal case, represented by an ideal city, the pitying part does not have a place. These three arguments hopefully make a good case for the pitying part’s satisfaction of 3).

Since pity satisfies these three conditions, we have reasons to believe that the first characterization successfully illustrates it in these three aspects, from which

¹⁷ This word can have the same sense as “πάθημα”: see LSJ s.v. “πάθος”. For the sense of “πάθημα”, see footnote 11. It should be noted that translators unanimously render this word as “experience” or “suffering”, which is quite plausible, though I think a case can be made for rendering it my way.

¹⁸ Thus, I cannot agree with G. R. F. Ferrari that “ἐπιεικής” should be rendered as “respectable” and that “the respectable fellow’s reaction is conditioned above all by his sense of propriety” (107). On the first point, while “ἐπιεικής” does have social or political connotations, Plato has Socrates use it almost exclusively for its moral sense. It is perhaps comparable to the concept of “君子” in classical Chinese philosophy. I prefer G. M. A. Grube’s rendering of it as “decent” or Paul Shorey’s as “the better sort”. On the second point, the fact that the decent man suppresses his grief more in the company of others only shows that spirit and reason work in tandem in this case, which is what we would expect from 440e, and it is not evidence for spirit’s rule within the soul. More importantly, when the decent man exercises pity and is thus corrupted at 606a, it is his *reason* which fails to see that pity on others will turn into self-pity, and not his spirit which fails to see this. Why, indeed, should spirit care about whether the tears are for the agent themselves or for a performance? Thus, I see no evidence in this part of the text for the claim that the decent man’s sense of propriety is what motivates his good conducts.

¹⁹ See Lear 208-215.

we would be justified to call it anti-reason. Now, from the first characterization, it appears that an important implication of calling a part anti-reason would be its separation from the three parts of the soul given in Book IV. However, this does not always follow, for an overgrowth of appetite and/or spirit too would satisfy these conditions, and can thus be called anti-reason. We cannot yet decidewhether Plato's arguments show him to be maintaining the tripartite theory in Book IV, or whether he is delineating a new part dedicated to the tragic experience. To determine this, we must examine whether the pitying part can be collapsed into spirit or appetite.²⁰ Since the pityingpart is responsible for the pleasure we experience in tragedies, figuring out the precise theory behind it is essential to understanding Plato's response to the problem of tragedy (606a).

VII. PITY, SPIRIT, AND APPETITE

To determine whether the pitying part can be identified with appetite or spirit, I will be working with two premises. The first is a version of what is by convention called Plato's principle of opposites, which, in Book X, is formulated as: "when there are opposite impulses ina human being in relation to the same thing at the same time, we say that there must be two elements in him" (604b). The second premise is that a part or an element of the soul cannot be divided against itself. By this, I mean that, while the appetitive desire can be divided into, for example, the desires for food and for drink, among many others, the appetitive desire cannot be divided into a desire which would satisfy the appetite, and a desire which would not satisfy the appetite, for, manifestly, the latter is no longer an appetitive desire.²¹

I will start with Hendrik Lorenz's useful characterizations of spirit and appetite. Appetiteincludes "the responses of the soul to unbalanced bodily conditions such as dehydration, and...desires for suitable processes which restore the body to its properly balanced condition". Spirit can be characterized as "the desire to distinguish oneself, typically by outdoing and outshining others, and to earn admiration and esteem by doing so" (*Analysis* 158-159). More, ofcourse, could be said, but these characterizations should be good enough for my purpose.²²

From here, it is quite obvious that pity cannot be collapsed into spirit. Socrates claims at 604a that a decent man fights his grief more in public when "he is seen by his equals." This means that, in public, towards the same object, grief, one part of the decent man desires to pity himself and mourn, while another part is afraid of the judgments from his peers and thus desires the opposite. According to the first premise, this means that two parts can be distinguished in thedecent man, and the latter fits with the characterization of spirit. According to the second premise, the other part cannot be a division of spirit, for it does not satisfy spirit. Thus, pity cannot be collapsed into spirit.

The case with appetite is more complicated. While it is easy to imagine a scenario where someone's grief grows so overwhelming that it interferes with their appetitive desires, this tendency would only resemble the case described at 485d: "when someone's appetites are strongly inclined in one direction, we

²⁰ I take it to be obvious that it cannot be collapsed into reason.

²¹ The second premise is similar to Terence Irwin's distinction: "Opposition 'to an appetite' may be simply an aversion to this appetite for this object; alternatively, it may be opposition to acting on appetite as such" (207).

²² See §6 for a characterization of pity.

surely know that they become more weakly inclined in the others.”²³ This does not exclude the possibility that the two desires come from the same part. For example, we surely see people who, because their passion for a particular subject has grown, begin to neglect other intellectual pursuits which they used to be passionate about.

But a determination through another route is possible: since the lower part is depicted as primarily a self-pitying part, it takes pleasure in experiencing and indulging in its own unhappiness (603e-604a, 605d). But Plato’s conception of happiness involves each part in the soul exercising its naturally suitable function, and to desire happiness is thus to desire this *healthy* constitution of the soul; properly speaking, then, only reason can desire happiness, since only reason has a view of the whole soul (352d-354a, 443c-444a, 441e).²⁴ That which finds pleasure in unhappiness would then derive satisfaction at beholding a sick soul in the person, whose constitution is arranged in such a way that it is as contrary to nature as possible (444c-e). But this part would then have a view to the whole soul—and it must, if what it desires is its own unhappiness. Since the appetite does not share reason’s view of the *whole* soul, this part cannot be the appetite.²⁵

One might say that the appetite takes pleasure in unhappiness too. For surely an appetitive desire could take pleasure in, say, excessive eating, which strengthens the appetite and thereby disrupts the natural order of the soul, and in this way, the individual becomes unhappier. In response, it is important to note that the satisfaction of an appetitive desire *simpliciter* has nothing to do with happiness or unhappiness—it has only to do with the satisfaction of desires whose objects are external things. It is only the satisfaction in conjunction with reason being overpowered that results in an unnatural ordering of the soul, but this ordering is not essential in any way to an appetitive desire. On the other hand, to give pleasure to self-pity, which is the part in question, is to be unhappy: the lack of order inside the soul is *essential* to its satisfaction, for the object of its desire is no external thing, but the soul itself.²⁶ But this can only mean that it has a view of the whole soul, and from this, the part must be distinguished from the appetite.

Thus, it has been shown that the part which takes pleasure in pitying is neither a part of the spirit nor of the appetite. Therefore, Plato’s psychological theory in Book X has developed into a quadripartition instead of the tripartition in Book IV. With this overarching general theory in place, we can understand the pleasure we experience during tragedies as peculiar to the pitying part, an irreducible and independent part of our psyche. This part, according to Plato, is

anti-reason, and hence by satisfying it and allowing it to rule in the soul we “put a bad constitution” in ourselves (605b). What remains to be said, now, is an account of how tragedies can target and affect this part of the soul.

²³ See, for example, Achilles’ sorrow for Patroclus ridding him of hunger: “Each man quickly made ready his evening meal, and then / they feasted, and no one’s heart lacked a fair share in the feasting... But Peleus’s son lay on the shoreline of the thunderous sea, / heavily sighing, with all his Myrmidons around him, / in a clear space, where the combers came crashing on the beach” (23.55-61).

²⁴ This line of argument is not altogether straightforward. For fuller accounts, see Irwin 244-256; Cooper 189-194.

²⁵ See Irwin 209-211.

²⁶ We should not take Plato too literally when he characterizes the part as having a “hunger (πεινῆρκός) for weeping and lamenting,” for he characterizes even reason’s desires with somatic terms (606a, Cooper 190). This point is in effect acknowledged by Moss though she still uses it as evidence that the part in question *can* be collapsed into appetite or spirit (*Appearances* 43).

VIII. TRAGIC TRANSPOSITION

When listening to a performance of Homer, according to Plato, even the best of us “give ourselves up to following it, *suffer along* with the hero, and take his sufferings seriously” (605d). How could we “suffer along” (συμπάσχοντες) with characters in a play when their misfortunes are distanced worlds away from us? Plato may give an answer if he claims that the pitying part not only takes pleasure in the agent’s own unhappiness, but enjoys grief *qua* grief, and thus would pity and suffer indiscriminately, even when the unhappiness does not belong to the agent. Does Plato claim that the pitying part has this characteristic?

He does by having Socrates say that, “the enjoyment of other people’s sufferings is inevitably transferred to one’s own, since, when pity is nourished and strengthened by the former, it is not easily suppressed in the case of one’s own sufferings” (606b). Surely, however, this “transference” would be impossible if the pitying part only engages in self-pity, or that the part of our soul responsible for sympathizing with others is different from the part which pities ourselves. Instead, just like the wine-lover who “loves every kind of wine and finds any excuse to enjoy it” (475a), the pitying part relishes any and all occasions where it can pity and mourn, either at someone else or at the agent themselves. It does not distinguish one’s own suffering and someone else’s, for its object is grief *tout court*.

This has been the final piece of Plato’s response to the problem of tragedy. Since the pitying part does not discriminate as to what its grief is about, it is easy for us to identify with what is at stake for the characters on stage and feel distressed when the characters are consumed by ruin. Yet since this distress gives pleasure to the pitying part, and since this part is usually suppressed (606a), we come out of a tragedy feeling relieved and liberated. The downside to this is that, once we allow the pitying part to have its fun, it becomes strengthened and more ready to feel pity for ourselves when we suffer from misfortunes;²⁷ and since the most important thing to do when misfortunes strike is to deliberate how to get back up (604c), the more we indulge in grieving the less we are doing the rational thing. Thus, what underlies the tragic pleasure is an anti-rational pitying part, and since we should be wary of gratifying this part lest it threatens reason’s reign, we should also be wary about the pleasures of tragedy.

There is, however, something more to be said. The problem of tragedy, as it has been developed up to this point, is starting to pose a significant threat to the central thesis of the *Republic*: namely, that the rule of reason and the resultant harmony constitute *justice* both in the individual and in the state. This threat is alluded to by Socrates’ famous remark about the “ancient quarrel between poetry and philosophy” (607b). To see what threat poetry poses, and thus gain a better understanding of the motive behind Plato’s response to the problem of tragedy, I will end this paper with a discussion of the inadequacy of Plato’s conception of justice.

²⁷ One of the general themes in Plato’s moral psychology is that a desire only grows stronger and more lawless through unrestrained satisfaction, but healthier and manageable through cultivation. See e.g. 588e, *Gorgias* 492e-494a, Irwin 117-118.

IX. THE PSYCHIC CONCEPTION OF JUSTICE

According to Plato, justice in the state is for each class to “do their own work” (433b), which translates to the individual level as someone who “does not allow any part of himself to do the work of another part or allow the various classes within him to meddle with each other” (443d); this is the person who achieves an inner harmony, in whom the rational part rules, “since it is really wise and exercises foresight on behalf of the whole soul” (441d). Plato’s conception of justice can be aptly termed *psychic justice*, because “it is not concerned with someone’s doing his own job on the outside. On the contrary, it is concerned with what is inside” (443b).

David Sachs famously declared Plato’s defense of justice a “fallacy”, because the conception of psychic justice is not what Thrasymachus, Glaucon, and Adeimantus have asked to be defended. However, while Glaucon and Adeimantus were indeed asking for a defense of justice, they also requested to know “what it is” (368c). It is then perfectly legitimate for Socrates to give a definition of justice based not on common opinion but on rational intuition alone—an intuition clarified through the city-soul analogy. There is nothing in the *Republic* in support of the contrary view, which takes common opinions to be the data that ground moral concepts. The times when Socrates compares his concept of justice to the common one are simply for confirmation, not for decision.²⁸ Thus, I do not think we should go as far as Sachs to claim that Plato fails for this reason.

However, even if Plato’s defense is not a fallacy, we might still question the adequacy of his conception of justice. We are justified to say that an important part of justice is its other-regarding character,²⁹ and this would not be bringing our own assumption into the text because Thrasymachus (343c) and Glaucon (358e-359b) too assume that justice is like this;³⁰ moreover, as Irwin shows, Plato’s own arguments agree that justice should have this other-regarding aspect, for “each part of the city does its proper work for the benefit of the whole city, and Plato clearly supposes that the rulers of the ideal city can be relied on to seek the good of the city” (259). Psychic justice appears to be decidedly lacking in this aspect, for it is concerned “with [the agent] himself, really, and the things that are his own” (443c).

Plato claims that the person who possesses psychic justice would also have the least to do with “temple robberies, thefts, and betrayals of friends” (443a). This amounts to I. M. Crombie’s suggestion that “by refusing to allow their passions and appetites to determine the direction of their lives, [the wise men] avoid the temptations which lead one man to trespass on the rights of another” (342).³¹ But surely it is at least *possible* that an occasion arises where one is absolutely required

²⁸ Sachs cites 442d in support of his thesis that “Plato thought that men who were just according to his conception of justice would pass the tests of ordinary morality” (47). But all that 442d tells us is that conformity to “the ordinary cases” would help *confirm* Socrates’ concept of justice. The fact that a thesis is confirmed if there is conformity to x only tells us that the lack of conformity to x does not confirm said thesis—surely it is a fallacy indeed to infer from this that the lack of conformity *disproves* the thesis.

²⁹ Gregory Vlastos observes that “δικαιοσύνη [justice] is the pre-eminently *social* virtue: it stands for right dealings between persons. And this is precisely what is missing in the Platonic definition, which purports to define a man’s justice in terms of the order which prevails within his psyche” (116).

³⁰ And also Aristotle; see *Eudemian Ethics* 1130a.

³¹ This is also in effect what Vlastos proposes as an emendation for Plato’s argument by suggesting that “the ultimate sources of wrongdoing—of injustice—are sensuality, cupidity, and vanity... [the first two] make up that monstrous beast of physical appetite which is ever ready to grow new heads in every direction...” (133).

to rob a temple so as to maintain the psychic order within themselves. If this should sound like an unlikely scenario, it becomes a lot more central to Plato's central argument in the Republic when we consider the famous descent of the philosophers back down into the cave to carry out the duties of "guard and care for the others" (519e). This is what Socrates proposes to say to them so that they would be persuaded to take up these duties:

When people like you come to be in other cities, they are justified in not sharing in the others' labors. After all, they have grown there spontaneously, against the will of the constitution in each of them. And when something grows of its own accord and owes no debt for its upbringing, it has justice on its side when it is not keen to *pay anyone for its upbringing*. But both for your own sakes and for that of the rest of the city, we have bred you to be leaders and kings in the hive, so to speak (520a-b).

Socrates' case here somehow relies on the Simonidean conception of justice introduced and left off in Book I, that "it is just to give to each what is owed to him" (331e),³² for the philosophers are expected to participate in political matters because they have received their education as a benefit from the city, and it is proper for them to repay it. Yet the philosophers would probably be justified in thinking that doing so would harm, or at least make it more difficult to maintain, their inner harmony, compared to simply philosophizing. The psychic conception of justice then would not be able to convince the philosophers to do what appears to be the just thing,³³ precisely because Plato's psychic conception of justice, contrary to the Simonidean one, does not account for the other-regarding aspect of justice.

Neither can the radical communalism in Book V forge a tight bond between the philosophers and the city: while the communalism might be effective to unite the auxiliary class, it is doubtful to what extent it can exert an influence on the philosophers, who desire the otherworldly Forms. Nothing, it can be argued, in the world is really *kin* to them, which is why they, out of everyone, are most able to accept misfortunes as they would "the fall of the dice" (604c). Yet if the study requisite for becoming a philosopher inevitably alienates them from worldly interests (485d), however, can they be counted on to take an interest in the city and political affairs? Why, in other words, should one believe that the philosopher has any regard for anyone other than themselves and the eternal Forms?

It is here that we see the true threat that tragedy poses to Plato's project. What binds the city together, and which therefore provides that other-regarding aspect of justice so difficult for the psychic conception to account for, is, ironically, "the sharing of pleasures and pains" (462b). But in order to share another's pain, we need a part in our soul that grieves indiscriminately, which doesn't only respond to one's own misfortunes but pities suffering *qua* suffering. This is, of course, none other than the anti-rational pitying part, the part which is targeted and developed by tragedies.

³² Irwin suggests that the Simonidean conception of justice and the definition Plato gives are not quite the same but "connected" (173). But, in fact, only in the Simonidean conception is the consideration for others a necessary component.

³³ Thus, Socrates alludes to the law forcing the philosophers to rule; see 519e, 346e-347e.

X. RIVALRY OF POETRY AND PHILOSOPHY

Plato, rightly I think, did not consider it possible for his psychic conception of justice to be complemented by a capacity for sympathy, which consists in and can be developed through grieving for another's suffering. For the rational response to misfortunes simply cannot dwell on them; the very *limit* of reason lies in its complete lack of comprehension that we may need to mourn for the sake of mourning (604b-c). Moreover, for reason to rule, clarity and precision must perpetuate throughout the soul. It is then unacceptable for it to mourn or grieve indiscriminately: it will, understandably, only grieve the good and beneficial grief.

I suspect that Plato tried to remedy the shortcoming in his conception of justice by legislating the radical communism among the guardians and by positing the Form of the Good; however, even if the latter is sufficient to give the philosopher an idea about what is good for the city and for everyone in it, it still gives them no motivation to carry it out, and the former policy fails to contribute to this end.³⁴

If these are right, then Plato was probably aware of this shortcoming in his conception of justice. He would also have been aware that, if one were to give up his psychic conception for this shortcoming, the natural alternative is offered by the poets, since their works develop our capacity to suffer together, and thus could account for the other-regarding aspect of justice.

Without making all of this quite explicit, Plato returns to the topic of poetry in Book X and, I think, gives his best to in turn point out its shortcomings. His point is that the sacrifice is too great in order to gain this capacity to sympathize: we would be putting up "a bad constitution in the soul" by feeding on "images that are far removed from the truth" and "gratifying the irrational part" (605b). His critique is thus less of an unprovoked attack, and more of moving his pieces one step ahead of his opponent.

XI. CONCLUSION

In sum, Plato answers the problem of tragedy by positing a pitying part within ourselves that takes pleasure in mourning and lamenting indiscriminately. I have gone beyond his answer itself in the final sections and questioned the motive behind his critique of poetry, not for the sake of idle speculation, but to give a reason for taking Book X seriously. Commentators have tended to treat this chapter either as a more or less unnecessary afterthought³⁵ or as deeply ironic.³⁶ Both of these ideas lead to the claim that, we somehow *help* Plato by considering the *Republic* to conclude in Book IX, that the work without its final chapter is a more coherent whole, and that we should somehow ignore a "doctrinal" reading of the arguments in Book X and instead focus on how the arguments are "framed". By pointing out the connection between tragedy and the shortcoming in Plato's

³⁴ Plato certainly does not claim that, after seeing the Good, the philosopher will think it is good for them to rule. If anything, the fact that a case has to be made to persuade them (with possible recourse to compulsion, see note 25) shows that Plato is not at all confident that they would want to participate in politics after seeing the Form of the Good.

³⁵ Annas, for example, comments that "Book 10 itself appears gratuitous and clumsy, and it is full of oddities. We can see why Plato thought it relevant to the rest of the *Republic*; but the level of philosophical argument and literary skill is much below the rest of the book" (333).

³⁶ See e.g. Halliwell, *Between* 155–207.

psychic conception of justice, I hope to have shown why neglecting Book X and its arguments would be a bad idea.

Should we, then, accept Plato's response to the problem of tragedy? On one hand, I think positing the pitying part is brilliant, and Plato rightly points to its deepest wish: to see and indulge in the agent's own unhappiness. A morality founded on this capacity to sympathize, and thus on the pitying part, according to Plato, is bound to produce *self*-pity in spades. On the other hand, I think he has missed some important merit about the capacity and desire to suffer together.³⁷

³⁷ What he misses, I think, manifests itself in his theories of love. For a very interesting account of a sort of narcissism necessary in Plato's conception of love, see Irwin 298-317.

BIBLIOGRAPHY

Adam, James. *The Republic of Plato: Edited with Critical Notes, Commentary and Appendices*. vol. 1, Cambridge University Press, 1905.

Annas, Julia. *An Introduction to Plato's Republic*. Clarendon Press, 1981.

Aristotle. *Eudemian Ethics*. Translated by Brad Inwood, Cambridge University Press, 2013.

Bobonich, Christopher. *Plato's Utopia Recast: His Later Ethics and Politics*. Oxford University Press, 2002.

Burnyeat, M. F. "Culture and Society in Plato's *Republic*." *Tanner Lectures on Human Values*, vol. 20, 1999, 215-324.

Cooper, John M. "Plato's Theory of Human Motivation." *History of Philosophy Quarterly*, vol. 1, no. 1, 1984, pp. 3-21.

Crombie, I. M. *An Examination of Plato's Doctrines*. Vol. 1, Routledge & Kegan Paul, 1962.

Du Bos, abbé (Jean-Baptiste). *Critical Reflections on Poetry, Painting and Music: With an Inquiry into the Rise and Progress of the Theatrical Entertainments of the Ancients*. Vol. 1, translated by Thomas Nugent, John Nourse, 1748.

Egil. "May 9: Bob Dylan—The Famous Horace Judson Interview in 1965 (video)." *All Dylan—a Bob Dylan Blog*, WordPress, 9 May 2017, alldylan.com/may-9-bob-dylan-the-famous-horace-judson-interview-in-1965-video/.

Feagin, Susan L. "The Pleasures of Tragedy." *American Philosophical Quarterly*, vol. 20, no. 1, 1983, pp. 95-104.

Ferrari, G. R. F. "The Three-Part Soul." *The Cambridge Companion to Plato's Republic*, Cambridge University Press, 2007, pp. 165-201.

Halliwel, S. *Plato: Republic 10*. Aris & Phillips, 1988.

———. *Between Ecstasy and Truth: Interpretations of Greek Poetics from Homer to Longinus*. Oxford University Press, 2015.

Homer. *The Iliad*. Translated by Peter Green, University of California Press, 2016.

Hume, David. "Of Tragedy." *Essays: Moral, Political, and Literary*, William Clowes and Sons, 1903, pp. 221-230.

Irwin, Terence. *Plato's Ethics*. Oxford University Press, 1995.

Lear, Jonathan. "Inside and Outside The Republic." *Phronesis*, vol. 37, no. 2, 1992, pp. 184-215.

Lorenz, Hendrik. "Plato on the Soul." *The Oxford Handbook of Plato*, 1st ed., edited by Gail Fine, Oxford University Press, 2008, pp. 243-266.

———. "The Analysis of the Soul in Plato's *Republic*." *The Blackwell Guide to Plato's Republic*, ed. Gerasimos Santas, Blackwell, 2006

Moss, Jessica. "What Is Imitative Poetry and Why Is It Bad?" *The Cambridge Companion to Plato's Republic*, Cambridge University Press, 2007, pp. 415–444.

———. "Appearances and Calculations: Plato's Division of the Soul." *Oxford Studies in Ancient Philosophy*, vol. 34, no. 1, 2008, pp. 35-67.

Nehamas, Alexander. "Plato on imitation and poetry in Republic 10." *Plato on Beauty, Wisdom, and the Arts*. Edited by J. M. E. Moravcsik and Philip Temko, Rowman & Littlefield, 1982.

Neill, Alex. "Tragedy." *The Routledge Companion to Aesthetics*, Edited by Berys Gaut and McIver Lopes, Routledge, 2002.

Plato. *Complete Works*. Edited by John M. Cooper, Hackett, 1997.

———. *Republic*. Translated by C. D. C. Reeve, Hackett, 2004.

Sachs, David. "A Fallacy in Plato's Republic." *The Philosophical Review*, vol. 72, no. 2, 1963, pp. 141–158.

Schlegel, August Wilhelm von. *A Course of Lectures on Dramatic Art and Literature*. Translated by John Black, 2nd ed., J. Templeman, 1840.

Vlastos, Gregory. *Platonic Studies*. Princeton University Press, 1973.

