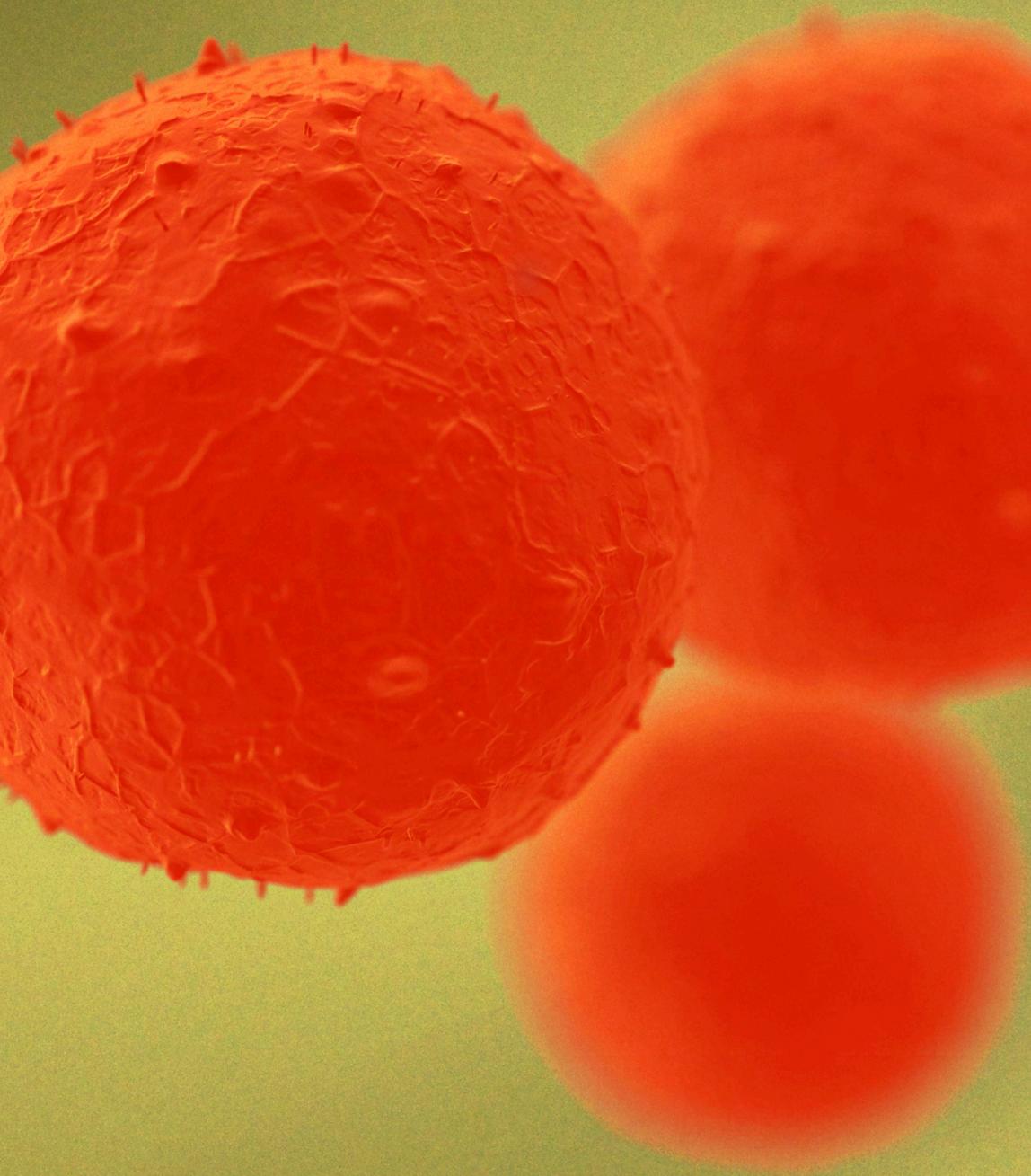


compos mentis

Undergraduate Journal of Cognition and Neuroethics
Volume 3, Issue 1



compos mentis

Student Editor

Thomas Mann

Production Editor

Zea Miller

Publication Details

Volume 3, Issue 1 was digitally published in May of 2015 from Flint, Michigan, under ISSN: 2330-0264.

© 2015 Center for Cognition and Neuroethics

The *Compos Mentis: Undergraduate Journal of Cognition and Neuroethics* is produced by the Center for Cognition and Neuroethics. For more on CCN or this journal, please visit cognethic.org.

Center for Cognition and Neuroethics
University of Michigan-Flint
Philosophy Department
544 French Hall
303 East Kearsley Street
Flint, MI 48502-1950

Introduction

Thank you for your interest in the publication. This is, I am pleased to say, the first issue in the third year of *compos mentis*. Our goal for this undergraduate publication has always been to provide a source in which undergraduates can, either for the first or hundredth time, submit or publish their work in a journal that focuses on a multidisciplinary approach. This undergraduate journal is unique in its commitment to remain at once focused and broad. Students from across the humanities and sciences have submitted their work for this issue and previously, and it lends itself to a unique perspective that is not bogged down using the methods of a single discipline. I also hope this journal gives many students the opportunity to read their peers' work and foster confidence in their own. This issue stems from the proceedings at the Third Annual Michigan Undergraduate Philosophy Conference.

Because of the interest we have garnered over the past three years, we continue to make *compos mentis* a bi-annual journal. Our review process begins in the winter, when we ask for the abstracts of papers to be submitted. All papers go through a blind review process, which includes faculty members as well as student philosophy club members. Acceptance into the conference and journal is based on the agreement of the blind review process, with emphasis placed on the opinions on the students. When all accepted papers have been finalized, including edits, and the review board agrees, we publish the issue.

I would like to thank those students of the philosophy club who have been very helpful with the review process. And thanks to the faculty advisers, Dr. Simon Cushing, Dr. Bénédicte Veillet, and especially Dr. Jami Anderson for their guidance and advice. I would also like to thank the Philosophy Department at the University of Michigan—Flint and to the Center for Cognition and Neuroethics. And, finally, thanks to our continually devoted Production Editor, Zea Miller.

This current issue of *compos mentis* features a variety of papers from students across the United States and Canada. We are excited and proud to offer these works.

Thomas Mann
May 2015
Flint, Michigan

Dissolving the “Other Person” Problem

Charles Dalrymple-Fraser

University of Toronto

ABSTRACT

Recent medical advances in detecting genetic dementia allow individuals diagnosed with prospective dementia to prepare for their futures as dementia. Typically, planning takes form in an advance directive, which allows prospective patients to detail their health care decisions in case of loss of capacity for consent. However, the validity of advance directives for dementia has recently come into question through what is known as the “other person” problem. The “other person” problem claims that advance directives should not be considered binding or valid in situations where the dementia patient seems to be a different person than the author of their advance directive: what right can one’s plan for oneself have to guide the care of another individual? Those seeking to defend the validity of an advance directive have mainly sought to establish a robust metaphysical theory of identity which can account for personal identity persistent through dementia onset. In light of the demonstrable failures of these approaches, I argue that there seem to be consistent moral intuitions already held in many applied ethical cases which, when mapped onto the case at hand, suggest that advance directives should be considered valid. In this way, this paper argues that it is a mistake to think that the continuity of personal identity is necessary to establish the validity of advance directives, and seeks to dissolve the “other person” problem.

KEYWORDS

Advance Directives, Alzheimer’s, Applied Ethics, Dementia, Ethics, Future Planning, Identity, Personal Identity

PUBLICATION DETAILS

A previous version of this paper was published as Dalrymple-Fraser, Charles. 2014. “Eliminating the ‘other person’ problem for advance directives.” *Mindful* 6: 68–91.

This paper has four sections. In section one, I introduce the “other person” problem for dementia. In section two, I indicate how our predominant intuitions in other applied ethical cases seem to map onto the case of dementia and advance directives, and suggests that validity may be constituted in the coherence of our current practices and intuitions. In section three, I respond to a few objections to the project at hand. Finally, I offer a brief conclusion in section four.

1. DEMENTIA AND THE “OTHER PERSON” PROBLEM

As a result of medical and technological advancements, more young individuals are being offered the opportunity to know, and the chance to prepare for, their possible futures as patients of dementia. One way for individuals to prepare for their futures as dementia patients is to draft an advance directive. Advance directives are legally protected documents which detail the wishes of a person toward their health care, in case of an event in which they have lost their capacity to make decisions. Advance directives are becoming increasingly legally—or “quasi-legally”—binding in many countries and can work to ensure that a patient’s wishes and autonomy are respected, even when they are incapable of decision making (Burnette and Heck 2012; Epstein 2007). Accordingly, advance directives offer a unique opportunity for future dementia patients, in that an advance directive allows them to determine their own health care without placing faith in and a burden on surrogate decision makers. However, with an increasing number of individuals turning to advance directives, there has been a corresponding increase in attention to whether advance directives ought to be considered valid for dementia patients.

In particular, concerns are raised as to whether the person receiving treatment in accordance with the advance directive—the moral patient—is necessarily the same person who details the advance directive—the moral agent. Indeed, radical change is a hallmark of dementia, with many patients suffering from: (i) memory impairment; (ii) changes in character; (iii) aphasia, apraxia, and agnosia; (iv) disturbances in executive functioning; and (v) significant impairments in social and occupation functioning (DeGrazia 1999). If it is possible that the moral agent is a numerically distinct person from the moral patient, then it becomes difficult to see whether advance directives for dementia patients should be considered valid.

This is the “other person” problem: the concern that an advance directive is invalid if there is no continuity of identity between the moral patient and the

moral agent. If the moral agent is not the same person as the moral patient of their advance directive, it is not clear that the agent should have the right or ability to make decisions for that patient, let alone that their decisions should be able to override the decisions of the patient. Yet, we retain an intuition that a person told that they will develop dementia should have a right to make decisions for what appears to be their future selves. Indeed, it seems odd that a future dementia patient can give power of attorney or otherwise have a surrogate decision maker, whose is able to make decisions on their behalf when they are suffering from dementia, but that a future person cannot themselves make decisions toward such future states. Granted, there is a distinction to be made between a surrogate decision maker and an advance directive, in terms of the ability of the former to react to different situations, but it is not clear that an advance directive is not capable of providing a rigorous anticipation of different events and changes in technology or condition.

Traditionally, philosophers have sought to resolve the apparent “other person” problem by attempting to identify a robust theory of personal identity which accounts for the intuition that there is some semblance of continuity of identity. However, it is not clear that any of these attempted solutions have resolved the issue. Indeed, an attempt to find continuity in physiological and psychological accounts of identity seem to fail to make any progress toward a resolution of the “other person” problem, precisely because they depend on the immutability of those features which suffer drastic change during dementia: the physiology of the brain, and the psychological characteristics and memories of the individuals. And, social accounts of identity which locate identity in the relations one holds to other people are strained by the changes in relationships between patients and their social networks: dramatic changes in patient character and memory, and the very strain of caring for a dementia patient all affect the relations which a dementia patient bears to other people. An extensive evaluation of the problems which theories of personhood face for dementia patients is beyond the scope of the present paper, and it is not my intent to discard the metaphysical project entirely. For, should an account of identity be constructed which can ground the validity of an advance directive, then that should sufficiently resolve the issue at hand.¹ Rather, I move to present a new account, which locates

1. Indeed, in later papers, I argue that an intersubjective account of narrative identity may be able to perform just this task. See Dalrymple-Fraser 2014, 2015a.

the validity of an advance directive as constituted in something other than personal identity, and hence dissolves rather than resolves the “other person” problem.

2. VALIDITY WITHOUT IDENTITY

In this section, I defend the claim that identity is not necessary for the validity of an advance directive by demonstrating that such an alternative account finds support in our common moral intuitions toward a number of other applied ethical cases. In particular, I will demonstrate that: (i) there are cases wherein we seem to have a moral right or obligation to make decisions for individuals who (a) are not yet in a position to make decisions for themselves, (b) cannot form moral or social contracts with us, and (c) are incapable of reversing our actions or the effects of our actions; and (ii) these intuitions map onto the concerns about dementia.

2.1 Conception

An interesting feature of the “other person” problem is that it concerns a person which may not yet exist. So, in prospectively considering the validity of an advance directive presently being signed, we are concerned with a possible future person.² There are a number of current ethical cases in which we seem to think we have moral rights or obligations to future persons, even if they do not yet exist or cannot form moral contracts with us. In this section, I look at such intuitions in the matter of conception. In §2.2, I turn briefly to future persons in the context of climate change.

An emerging theme in the discourse surrounding conception is the intuition that we may have obligations not to reproduce when it is for the betterment of possible people. In particular, Laura Purdy (1996) argues that persons who suffer from Huntington’s Disease ought not to reproduce, given the high probability that their offspring would suffer from Huntington’s Disease. Specifically, she argues that a life ailed by Huntington’s Disease can be so terrible that it is morally reprehensible to bring into existence a person who is extremely likely to be

2. Philosophers, particularly Derek Parfit (cf. Parfit, 1976), tend to be careful to distinguish between possible people and future people. Here, I will take upon the practice of referring to the moral patients here as possible future persons. For, if they do exist, then their existence is not contingent on our present actions (unless a cure is found); but their existence is not certain. I will keep the terminological distinction intact for the other discussions which follow.

plagued by that disease; this position is motivated by our contemporary knowledge of the disease, its symptoms, and its 50% inheritability. Furthermore, given a context where alternative means of raising a family are available, such as adoption or surrogacy, the risk of conceiving an individual with a tragic inheritable disease seems unjustified. If Purdy's intuitions are met, then we have a case of moral obligations to numerically distinct persons. Moreover, the nature of the relation is such that the moral patients (the possible children) are incapable of consent and of entering into a moral contract with the moral agents (the potential parents), and are unable to reverse the decisions made by the moral agents. Still, the prospective parents have a right and obligation to make decisions affection those possible children.

However, Purdy's argument stands against other philosophers' claims that existence is inherently better than nonexistence, and it is not immediately clear that her position is the most common one, even if it is very intuitive. It is not the purpose of this paper to decide such disputes. Yet, it is worth noting that many of those philosophers who hold that existence is categorically better than nonexistence would not be averse to the claim that, should it be proven better in a case for a possible person not to exist than to exist, then a moral agent would have an obligation not to reproduce. Indeed, the difference between the positions seems to fall to whether there does exist such a situation wherein it is better not to exist. Hence, it seems that the salient matter stands, namely that we can have certain moral obligations to possible future people, and rights to make decisions concerning them.

2.2 Climate change

The motivation for thinking that we have moral obligations to future generations in the case of climate change seems to consist in the same principle which drives considerations of obligations not to conceive: the intuitive principle that we should prevent or mitigate harm to others. In this manner, climate change can be considered akin to a hereditary disease like Huntington's Disease: the severe predicted effects of climate change will negatively impact the quality of life in a very extreme manner for those future generations, and given the extreme likelihood that climate change has anthropogenically causes, it appears even more of a 'hereditary' risk than does Huntington's Disease (IPCC 2013, 15).³ Here,

3. I defend this pathologization argument in further detail elsewhere. See Dalrymple-Fraser, 2015b.

then, we sustain an intuition that we have a moral right, if not an obligation, to make decisions which affect future generations and future persons, which are numerically distinct from ourselves. Furthermore, future generations are incapable of forming moral contracts, given that they do not yet exist when the relevant decisions need to be made, and are incapable of reversing the effects of our decisions (IPCC 2013, 25–27).

Accordingly, we can draw another parallel to the matter of advance directives. It seems that common intuitions about climate change ethics fit uniformly in line with what would be required of an account of validity for advance directives, without personal identity.

2.3 Nonhuman animals

Perhaps more common than discourse on environmental ethics as a whole, is the discourse on whether we ought to extend morality to animals. Again, one need not delve into the literature to meet the common intuition that we ought to treat animals with a particular moral regard. Such an attitude is reflected in the common treatment of pets, and the disgust which meets reports of animal cruelty. Regardless of the philosophical approach to animal ethics (e.g., consequentialist, deontological),⁴ there tends to be strong agreement with the claim that we have obligations to treat nonhuman animals in a particular moral regard, regardless of whether that treatment is exactly equal to the treatment of human animals. The typical force of this intuition is such that I will not examine it in detail, but demonstrate how it maps onto our present case of advance directives. For, it should be obvious that nonhuman animals are distinct from us, that they are incapable of the communication required to form a moral contract, and that they are largely incapable of reversing our decisions: our choices to conserve or display animals, to breed them or let them expire, to log and depreciate their habitats, are actions which nonhuman animals are largely incapable of offsetting. Again, we find an intuitive fit between our current ethical intuitions and practices, and the case of advance directives.

Granted, not everyone holds that we have moral obligations to treat nonhuman animals in certain ways. Yet, most of these objectors hold that we have rights to make decisions about the treatment of animals—to zoo them, to hunt them, to raise them for the sole purpose of consumption, and so forth. And, even

4. See, as exemplary, Singer (1974) and Regan (1983), respectively.

if we have a mere right to make such decisions, again, it is a right to treat certain numerically distinct entities in particular ways, entities incapable of consent and of reversing our decisions. Indeed, both sides on the animal ethics debate seem to allow room for the sort of intuitions and practices we wish to bring to bear on dementia.

2.4 Guardians and minors

Finally, let us briefly consider parenthood and guardianship over minors. This case is useful in that it regards a widely visible practice and calls upon strong intuitions, as with the case of nonhuman animals. Indeed, it is widely regarded that children are rarely capable of making their own informed decisions. In these circumstances, it is their parent or legal guardian who provides consent, and who makes decisions which concern those children. It does happen on occasion that a parent or guardian makes a decision which is deemed wrong for the child, but it is not contested that parents have a right—if not an obligation—to make decisions for their children.

This strong intuition maps readily to the matter of dementia. It is widely regarded that those suffering from severe dementia are frequently incapable of making their own decisions; indeed, it is because of this diminished capacity that questions regarding advance directives are raised at all. Furthermore, we wish to attend to the possibility that the moral patients of our advance directives are numerically distinct persons, whence the “other person” problem arises, and children are clearly nonidentical with their parents. Moreover, children are also widely incapable of reversing the decisions set out by their parents or guardians, as are patients of dementia acted upon through the advance directive. In these manners, parent–child relationships map readily to our advance directive case.

2.5 Interim Summary

The cases considered here are not meant to be exhaustive nor conclusive. However, there is clear precedent for accepting the possibility of a theory of the validity of advance directives which does not rely on the continuity of personal identity. Indeed, there is a wealth of intuitions supporting the claim that we may make decisions for other persons and entities, and a number of these cases find reasonable congruency with the case at hand. Hence, we may seek to establish

the validity of advance directives in the coherency of our currently adopted systems of intuitions and practices.

However, it is worth remembering that the purpose of this paper is not to advance a strong theory of what this validity may precisely consist in. Rather, its purpose is to demonstrate that a survey of common intuitions shows a consistency within those intuitions which may inform such a theory. For example, we might find that the validity of an advance directive can be constructed from the particular strength or closeness of the relationship between the moral agent and the moral patient, just as it seems that the particular closeness of relation between a parent and their child grants the right of the former to make decisions concerning the latter. However, an examination of such speculations is beyond the scope of this paper. Still, the cases presented suggest the possibility that a theory may be constructed which, in accounting for these above intuitions, accounts also for the validity of advance directives in cases of dementia.

3. OBJECTIONS CONSIDERED

In showing that common intuitions about many applied ethical cases map uniformly onto the debate over the validity of advance directives for dementia patients, this paper has paved the way for a theory which can account for intuitions in favour of the validity of advance directives, even in circumstances without promise of continuous personal identity. However, it is worth noting that a few objections may be raised against this move, even before such a theory has been fully articulated. In this section, I briefly examine three such possible objections, and demonstrate how they fail to undermine the current project.

3.1 Acting on Intuitions

A first objection may be raised, that it is not clear that moral intuitions ought to guide our moral theories or practices. Indeed, our intuitions may be wrong or poorly grounded. Certainly, each case in the above section is a matter of philosophical contention: debate rages over whether we have obligations to animals or to future and possible generations. Accordingly, it is not clear that these intuitions are sufficient to motivate an account of validity for advance directives.

However, predominant moral intuitions do seem to play a role in informing policy and practices. It is a mistake to treat the issue at hand as a debate to be

merely relinquished to the philosophical armchair. Rather, the project of reconciling advance directives with the concern of non-continuity of identity is one which plagues our current affairs. Ethics in practice requires decisions to be made, and policies to be enacted, and there is a wealth of evidence to support the claim that these decisions are made with consideration to such moral intuition: one need only to look at the increasing number of abortion clinics, to movements against climate change, and progress toward an animal ethics, in order to identify that our moral practices do not wait for theoretical unanimity, and that do not they take place in an intuitive vacuum. It may be the case that our intuitions are wrong, but this should not prevent us from making progress in practice and policy at present, where such progress should be made in an attempt to solve the problems at hand. Rather, we should recognize and contend that policies may be corrected as we gain more moral data or theoretical constructs, just as we tend to do for other practices which are updated with the changes in scientific knowledge. In this regard, the objection seems to miss the applied matter at hand, focusing too strongly on the need for a coherent theoretical system.

3.2 Transferring intuitions

A further objection may be raised that it is not necessarily the case that moral intuitions carry across seemingly similar cases. That is, even if we permit moral intuitions to play a role in practice and policy, it is not necessarily evident that the moral intuitions in the above cases do carry precisely to the matter of advance directives. Indeed, examinations of trolley-like cases have demonstrated apparent contradictions in our intuitions. It is not clear that our moral intuitions about other applied ethical cases should carry into the moral matter of advance directives and dementia patients. That is, there may be a subtle and morally relevant difference between the cases we have considered, which renders our intuitive bridging invalid.

The response to this objection is brief: the burden of proof falls to the interlocutor to demonstrate wherein this difference consists. For, it seems intuitive that a future dementia patient has a right to make decisions for what appears to be their self, even if it turns out to be the case that the dementia patient in the future is a different person. Unlike other cases in conflict, there is not a strong difference of intuition which suggests a morally relevant difference between the cases presented in section three and with advance directives for dementia

patients. Accordingly, it falls to the interlocutor to demonstrate either where our intuitions separate, or wherein a morally relevant difference consists.

3.3 The “other” person

Following the above objection, one might gesture that there is a critical difference between the dementia case and the other cases considered. The concern with the “other person” problem is that the decisions we make for ourselves ought not to be applied to other dissenting parties. That is, the problem is a problem in particular because we are choosing to act for ourselves, not for others as we do with children and future generations.

The tentative response here comes in two parts. The first is that we can conceive of one filling out an advance directive for the dementia patient, where one’s becoming the dementia patient is but a high likelihood. With a lack of phenomenological understanding of the disorder, many prospective patients are uncertain of how much of “them” remains through dementia onset. Accordingly, so long as one anticipates that the dementia patient may not be them, then if the author of the advance directive and the dementia patient are sufficiently related—an articulation of which is beyond this groundwork, but an intuition of which should carry—the validity ought to stand. A second brief note is that few people are perhaps best able to make decisions for dementia patients than those who are related to those patients through the anticipation of being those patients. We largely lack a phenomenology of dementia, and it certainly seems reasonable that a prospective dementia patient be given priority in decision making over a more detached surrogate decision maker. Ultimately, an adequate rejection of this objection will depend in part upon the theory established to account for our coherent intuitions and practices. Still, even with these few notes, it seems that the objection raised can be similarly dissolved in practice.

4. CONCLUSION

This paper advanced the position that validity need not necessarily consist in personal identity, and demonstrated how this position found support in pre-established moral intuitions across an applied ethical field. In doing so, we marked a novel approach to dealing with the “other person” problem, which does not rely on personal identity but keeps advance directives intact.

The content of this paper should present without surprise, for it does nothing more than to dissect the cases toward which we have ready intuitions. It should be plainly evident that it is consistent to hold a theory regarding the validity of an advance directive, without appeal to identity, given that we share these intuitions and practices in other applied ethical debates. Yet, despite their intuitive ease, these arguments are ultimately valuable as they may drive common intuitions back toward acceptance of advance directives for dementia patients, where physicians and families today are often faced with uncertainty. It is in this regard that we may rekindle optimism about the futures of future dementia patients.

WORKS CITED

- Burnette, Angela T., and Hannah Heck. 2012. "Advance Directives, Dementia, and Alzheimer's Disease." *Health Care Law Monthly* 11: 2–8.
- Dalrymple-Fraser, Charles. 2014. "Telling Stories: Narrative Identity in Dementia Pathographies." Paper presented at the University of Toronto undergraduate philosophy conference, Toronto, Ontario, December 3.
- Dalrymple-Fraser, Charles. 2015a. "From Autobiography to Biography: Narrative Identity at the Limits of Authorship." Paper presented at *Re-member: Memory, Narrative, and Identity*, Paris, France, April 10–11.
- Dalrymple-Fraser, Charles. 2015b. "The Conception Argument for Climate Change Enaction." Paper presented at the 2015 meeting of the South Carolina Society for Philosophy, Spartanburg, South Carolina, March 27–28.
- DeGrazia, David. 1999. "Advance Directives, Dementia, and 'The Someone Else Problem'." *Bioethics* 13 (5): 373–391.
- Epstein, Miran. 2007. "Legitimizing the Shameful: End-of-life Ethics and the Political Economy of Death." *Bioethics* 21: 23–31.
- Intergovernmental Panel on Climate Change [IPCC]. 2013. "Summary for Policymakers." *Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Switzerland: IPCC. Retrieved November 30, 2013. http://www.climatechange2013.org/images/uploads/WGI_AR5_SPM_brochure.pdf.

compos mentis

- Parfit, Derek. 1976/2006. "Rights, Interests, and Possible People." In *Bioethics: An Anthology*, edited by Helga Kuhse and Peter Singer, 108–112. MA: Blackwell.
- Purdy, Laura. 1996/2006. "Genetics and Reproductive Risk: Can Having Children Be Immoral?" In *Bioethics: An Anthology*, edited by Helga Kuhse and Peter Singer, 115–121. MA: Blackwell.
- Regan, Tom. 1983/1985. "The Case for Animal Rights." In *In Defense of Animals*, edited by Peter Singer, 13–26. Oxford: Blackwell.
- Singer, Peter. 1974/2006. "All Animals are Equal." In *Bioethics: An Anthology*, edited by Helga Kuhse and Peter Singer, 568–577. MA: Blackwell.

compos mentis

The Philanthropy Machine: Neuroethics and Empathy

Abigail E. DeHart

Grand Valley State University

Luke S. DeHart

Grand Valley State University

ABSTRACT

Imagine the Philanthropy Machine: mathematically complex and capable of making perfectly logical evaluations while using all available information to aid in decision-making processes. The Philanthropy Machine is so powerful that the humanitarian efforts of both agencies and individuals worldwide utilize it for maximizing the efficacy of their giving. It is high-powered enough to immediately identify the greatest human need and allocate the funds necessary. The Philanthropy Machine would presumably eliminate the "Identifiable Victim Effect" (I.V.E.), which refers to the tendency of humans to donate to a specific and identifiable person as opposed to a large group in need. But, what may be lost when a direct link is removed between donor and cause? Neuroethicists have taken up these questions of empathy and rationality when applied to the I.V.E., and whether empathy prevents logical philanthropic decisions. In this paper I will argue that, in the field of neuroethics, moral sentimentalism should be taken more seriously. Moral sentimentalist philosophers like David Hume and Adam Smith provided careful work on the analysis of empathy (what they called "sympathy"). Empathy, as they conceived it, would involve an attempt to understand the emotional and logical rationale behind every decision.

KEYWORDS

History, Moral Sentimentalism, Adam Smith, Altruism, Neuroethics, Empathy, Ethics

INTRODUCTION

Imagine an incredible machine: mathematically complex and capable of making perfectly logical evaluations while simultaneously using all available information to aid in decision-making processes. Let's call this machine the Philanthropy Machine. The Philanthropy Machine is so powerful that the humanitarian efforts of both agencies and individuals worldwide utilize it for maximizing the efficacy of their giving. It is high-powered enough to immediately identify the greatest human need and allocate the funds necessary to alleviate that need. With powerful algorithms that allow for near-instantaneous machine learning, the Philanthropy Machine might allocate funds to an emergency surgery in Africa, hurricane relief in India, or a nutritious meal to combat a child's hunger in America. It would be a step beyond utilitarian calculus, because the purpose would not chiefly be to help the greatest number, but rather anyone at all with the greatest need. The Philanthropy Machine will perfectly eliminate the bias inherent in charitable giving, bias that has the potential to disproportionately allocate resources. The question is then, if the Philanthropy Machine existed what may be lost when a direct link is removed between donor and cause? ¹

In this paper, we aim to assess what may be lost if something like the Philanthropy Machine was adopted as the primary method of allocating monetary donations. In so doing, we will draw on work of neuroscientists on the associated questions of empathy and rationality when applied to philanthropy, assessing whether empathy prevents logical philanthropic decisions. To this end, we will argue that, in the field of neuroethics, moral sentimentalism should be taken more seriously. Moral sentimentalist philosophers like David Hume and Adam Smith provided careful work on the analysis of empathy (what they called "sympathy"). Empathy, as they conceived it, would involve an attempt to understand the emotional and logical rationale behind every decision.

1. We offer a big thanks to the participants of the 2015 Michigan Undergraduate Philosophy Conference who provided their valuable insight, assessment, and reactions to this concept of the Philanthropy Machine. While there is much this imaginary machine cannot yet do (like account for donations of time, as one commentator pointed out), their conversation convinced me that there is much to be gained through a theoretical discussion, which can help us talk about and philosophically assess the value of philanthropy and empathy.

THE IDENTIFIABLE VICTIM EFFECT

It will first be useful to establish that there actually is an inherent bias in charitable giving. We need to look no further than the Identifiable Victim Effect, a phenomenon first classified by psychologists that demonstrates our tendency to give more to specific individuals in need rather than large groups, even if the need is much less severe in the case of the individual. It can lead to donors giving funds disproportionately to one cause over an equally and often times more pressing need. The most popular illustration of the Identifiable Victim Effect is the 1987 case of the 18-month-old child, Jessica McClure, who spent 58 hours trapped in a well. In addition to the remarkable rescue effort, in the months that followed, the McClures received over \$700,000 in donations in a spectacular outpouring of support from the nation. By comparison at the time, these same resources could have been used on preventable health care to save hundreds of children's lives (Jenni, Loewenstein, 1997). Cases like this depict the power of human empathy to address societal problems, but they also raise worrisome questions about the negative effects empathy may have on the efficacy of our charitable giving.

The Identifiable Victim Effect has been well documented, but we wish to turn the focus not to its existence but rather to considering such instances where our own empathy may be a hindrance as we reason through our charitable giving. We will explore how the Philanthropy Machine could factor in to these instances. For conceptual ease, let us consider the role of empathy in philanthropy as being located in two chronologically distinct stages: Stage 1 is the first moment of contact with a cause or charity and involves recognizing a need; Stage 2 is the decision to donate, which involves determining the amount of money to be given. Studies have shown it is in Stage 2 that empathy becomes an issue, as is demonstrated by the Identifiable Victim Effect. In one such study, investigators challenged the rational decision-making of their subjects by presenting them with certain scenarios that would invoke this Identifiable Victim Effect by varying the singularity of the victims. Subjects were presented with a single victim versus a group of eight individuals and were asked to give money to either over the course of several trials. Researchers found that when both groups remained unidentified, the subjects gave nearly equally to each, but when the single victim was identified by variables like age, picture, and name, this individual elicited much more contribution than an unidentified individual and group of eight. This same

identifying information, however, did not increase the monetary amount awarded to the group of eight. The subjects also reported that they felt much more distress when the individual victim was identified than in any other trial (Kogut, Ritov, 2005). Because the money could go further in helping the eight victims as opposed to the one, it would appear the more rational decision to give money to the eight. It is true that subjects still donated some money to both groups and we would argue that this is because the Stage 1 of empathy in philanthropic giving was activated—the subjects recognized the need of both groups. We propose that the discrepancy occurs in Stage 2, because empathy seems to have interfered with what could otherwise be a calculated decision regarding which group needed more money.

We also have some growing evidence of what happens in the brain when faced with decisions to donate. In one study from 2013, researchers looked at the neurological response as subjects were presented with a series of options for donation. In the first trials, subjects were only shown silhouettes chosen to represent potential donees and researchers tracked the neurological response. Researchers compared this with the response to that same group as represented by photographs and found an increased positive arousal in the brain. They also found that they could predict increased donations by tracking these neural correlates when subjects look at the photographs of potential donees (Genevsky, et. al.). Together these studies and others like them show the powerful affective response we have towards identified victims. It also indicates a difficulty in that statistical victims of misfortune can fail to bring about the same level of response as an identified victim—even if the larger group requires more assistance.

Given all of the research that has gone into the Identifiable Victim Effect, one may propose that this difficulty could be assuaged by educating a wider audience in order to encourage people to think more rationally about their philanthropic decisions. Another study was conducted with this very idea in mind, but it showed that even when subjects are informed about the Identifiable Victim Effect, it does nothing to increase the amount of money given to the statistical majority. Distressingly, however, it does drastically decrease the amount given to the identified victim (Small, Loewenstein, Slovic, 2007). The study also found that activating rational analysis can actually decrease donations in general. The experiment “primed” research subjects by having them complete simple arithmetic calculations and then decide how much to donate. It seems as though

exercising the rational brain can repress the intuitive moral feelings that arise which compel us to help one another.

These apparent irrationalities of our brains can cause large problems. There is disconnect that happens between our intentions and the decision-making process that leads to the final resting place of our philanthropic efforts. But, this raises the question of whether or not emotional responses should be removed from decision-making processes all together. We aim to assess what may be lost if something like the Philanthropy Machine was to be adopted as the primary method of allocating charitable donations.

AN EIGHTEENTH-CENTURY APPROACH TO NEUROETHICS

Neuroethicists have taken up these questions of empathy and rationality more recently, but connection between neuroscience and philosophy has typically been dominated by philosophy of mind and questions of free will. We argue that in the field of neuroethics, moral sentimentalism should be taken more seriously as it can provide a way to frame further inquiry into the role of empathy within philanthropy. We will explore moral sentimentalism through the Scottish Enlightenment tradition, as philosophers like Adam Smith and David Hume were among the first moral sentimentalists. These writers considered what they did to be “moral science”, which is demonstrated by the fact that they focused intensely on what the human was and did rather than what he or she should be. The tradition’s interest revolved around people: what motivated them, what brought them together, and what made society fairer. These questions were so all-encompassing that it would be difficult to house under any distinct discipline today, but this serves as a foremost example of the line of inquiry that suitably struggles with the ethical complexity of societal issues. The Scottish Enlightenment tradition had its own brand of scholarship, distinct from the larger concurrent Enlightenment tradition and provides a nuanced exploration of “empathy” and its role in fields ranging from economics to astronomy.

Though the roots of moral sentimentalism in ethical theory dates back to the early eighteenth century, it has largely been neglected in the field of ethics because it is now often regarded as in the realm of social science. But this “moral science” was more than an empirical exercise. We will focus specifically on Adam Smith and his work on “sympathy” (what we can most nearly translate as modern

day conceptions of “empathy”² and how it affected ethical and economic decision-making. It will examine Smith’s thoughts on globalization and his worries over a weak sort of benevolence toward those unknown to us that can occur over long distances. We believe that these moral philosophers and more specifically Adam Smith’s work *The Theory of Moral Sentiments* provide the framework to delve deeper into the role of empathy in philanthropy.

Adam Smith supposed that an operational system of morals was partially based on its capacity to account for a good theory of fellow-feeling. He used the term “fellow-feeling” interchangeably with “sympathy”, but empathy is probably the closest equivalent to what we now think of as the mechanism fulfilling this function. The key for Smith was the imaginary place changing that one does in order to feel sympathy. His theory of the impartial spectator used the role of individual imagination, enabling a person to enter into the imagined situation of another (Smith, 1790). So, according to his theory, one does not merely consider how it would feel to be a beggar on the streets, she truly envisions how it is to be that person. She imagines actually switching persons and characters, and to Smith, the emotions that arise from this place switching are not in the least bit selfish (Smith, 1790).

This place changing through the impartial spectator would ideally result in a nearly full switch with the other person, and so you would actually feel what it is like to be that other person in pain, thereby coming to a better understanding of how it affects them. To illustrate this, consider watching a child and an adult walk into the dentist to get a filling. Both are overtly experiencing the same phenomena, but were we to attempt to put ourselves in the place of the child, we might realize that any pain felt is also mixed with fear of the unknown and perhaps even doubt in the beneficial outcome of the action taking place. Through this place changing, we have come to feel via empathy for both the adult and the child, but might only be compelled to take action and comfort the child. Likewise, perhaps we find out that the adult had some traumatic experience early on with a dentist that makes every subsequent trip into a lived nightmare. It is not until we learn more through switching places as much as possible that we can respond in a way that might be the most appropriate.

2. While it is not within the scope of this paper to explain this move from “sympathy” to “empathy”, see Lou Agosta’s longer and more nuanced treatment of this in his book *Empathy in the Context of Philosophy*.

Smithian sympathy first involves comprehending one's own feelings in order to extend this emotional data to someone else. For Smith, sympathy is what also enables us to come to an understanding of what another feels like at a certain time or in a certain situation. It is not, strictly, "feeling bad" for a person because one can also experience feelings of happiness via sympathy for someone in a good situation. Smith explicitly distinguishes between sympathy and compassion, but he acknowledges a traditional overlap between the two noting, however, the generalization of sympathy:

Pity and compassion are words appropriated to signify our fellow-feeling with the sorrow of others, sympathy, though its meaning was, perhaps, originally the same, may now, however, without much impropriety, be made use of to denote our fellow-feeling with any passion whatever (1759).

Smith's account of sympathy takes in human circumstances. His emphasis on sympathy leads Smith to say that the impartial spectator can sympathize without having the same feeling as the agent, which he illustrates with two particularly outstanding examples: one in which the agent is dead and the other, the agent is mentally ill. Presumably, we would feel sad or upset via sympathy when we switch places, even if these two agents do not at the time. For Smith, sympathy is not some separate reactive effect that occurs when witnessing the pain and suffering of another individual.

Smith was also careful to emphasize the need to put oneself in the situation of another in order to most fully achieve the change of situation that gives rise to our sympathy. He is worth quoting at length here:

In all such cases, that there *may be some correspondence of sentiments* between the spectator and the person principally concerned, the spectator must, first of all, endeavour, as much as he can, to put himself in the situation of the other, and to bring home to himself every little circumstance of distress which can possibly occur to the sufferer. He must adopt the whole case of his companion with all its minutest incidents; and strive to render as perfect as possible, that imaginary change of situation upon which his sympathy is founded (Smith, 1759, emphasis my own).

The goal that Smith points out is to achieve some correspondence of sentiments—to actually feel what the other is feeling as derived from their own experience. For Smith, sympathy was not a motive for moral action (and not selfish at all), and he was clear about this.

Smith was very skeptical of our ability to match good intentions with good effects. In a recent book by Fonna Forman-Barzilai, she offers an analysis of Smith's circles of sympathy—a metaphor taken from the Stoic philosopher Hierocles to describe the natural phenomena of coming to care for someone else. Forman applies Smith's theory to the idea of cosmopolitanism and his analysis of sympathy vis-à-vis space (Forman-Barzilai, 2010). Smith was clear that the farther we are from someone, the more difficult it becomes to change places with him or her using the impartial spectator. Though we can presumably come to closer understandings of those we love and even those within our community, it becomes difficult when attempting to know what it is like to be someone living on the opposite side of the world. Smith had in his sights here examples like Britain's attempts at charity in colonial India. He was outspoken against the East India Company as well as the concurrent attempts to colonize the lands. Smith pointed out that it is more likely that our good intentions would be distorted at such a distances.

We can see why Smith was skeptical that human benevolence could properly reach across distances because even today international charity attempts fall far from their intended mark and can even leave the benefactors worse off than before. This is no reason to abandon international aid—in today's global economy, that option is near impossible—and even in the eighteenth-century, Smith did not think it was impossible for us to sympathize with those across the world. For him, however, to properly sympathize involved crucial knowledge component, which entails the ability to take on as much relevant information as possible and discover what it is like to be the other. Without this knowledge component, humans were only capable of weak forms of benevolence, which means that good intentions could not meet their mark of alleviating suffering or trouble. Still, Smith believed that it is very difficult for us to know our true intentions and to understand why we are motivated to act in some cases and not others. This is why the idea of the Philanthropy Machine becomes intriguing. This machine would ensure that even if the cause of giving were less than rational, the effect would still be good. We

could still be moved to give by Stage 1 of empathy, but Stage 2 would be taken care of by this machine, ensuring that the greatest needs are well met.

THE PHILANTHROPY MACHINE RECONSIDERED

The Philanthropy Machine necessarily removes the direct ties between donor and recipient, and by doing so this changes a part of the giving equation by affecting empathy's role in Stage 2 of philanthropic giving. The machine will take nothing away from the initial (Stage 1) feelings of empathy because we can still be empathetically moved to identify with a perceived need. However, Stage 2 will not be involved because when one donates to the Philanthropy Machine, the money will be distributed in the most efficient and effective way and this will not involve any further decision-making by the donor. The donor will still be doing good (indeed, the donor is guaranteed to be doing the most good), but he is, in some sense, separated from the efficacy of his giving.

Still, though the idea of loss usually denotes something negative, perhaps this separation from the efficacy of your donation is a justifiable loss if it can ensure successful results. Perhaps, in order to ensure people still donate, clever forms of marketing can be used to elicit a response similar to the Identifiable Victim Effect in order to inspire people to donate to the Philanthropy Machine. This will only become more possible because neuroscientific discovery has and will continue to provide empirical findings vis-à-vis how the brain reacts when faced with decisions to donate. In terms of any "loss", we wish to turn back to the discussion of moral sentimentalism in order to highlight one specific aspect of Smith's impartial spectator.

The study of morality for Smith started with an exploration of human nature, adhering to naturalism and interpreting it through history. A resounding point throughout most of Adam Smith's work was in the human desire to know and imagine when we don't know. While we want to direct altruistic tendencies toward better ends, it should also force us to take a step back and consider the descriptive implications that come from exploring empathy and altruism. Humans have an immense desire to know more. Smith observed that humans are curious about one another and that there is an entirely natural bent to sympathize with another. We naturally want our feelings to resonate with those we are spectating because we are curious to know why they do what they do. The more we can endeavor to place ourselves in the situation of another, the better sense we will

have of the propriety of the actions of this person. This will also allow us to act in ways that would be more fully empathetic.

As researchers interested in philanthropy (be it in the business or non-profit realm), this aspect of human nature should not be forgotten: that we are empathetic beings. There is something—and neuroscientific research can shed further light on what, exactly, it is—that wants to align sentiment with sentiment. Whether it is curiosity inherent in human nature, an evolutionary phenomenon, or something else entirely, we want to know more about “the other.” The more we find out about others, the more responsive we are, and as the Identifiable Victim Effect demonstrates, this does not always result in the greatest good for the greatest number; however, it is important not to eliminate empathy from the decision-making process even if it can lead to “irrational decisions.” While we seek to design better, more sustainable methods of eradicating poverty and improving lives, the importance of empathy to this process is irreplaceable. True empathy involves an attempt to understand the emotional *and* logical rationale behind every decision, which serves as invaluable information in decision-making.

Imagining innovations like the Philanthropy Machine might provide us temporary solutions, but attempting a full account of human nature and all of its complexity may involve some patience along the way as we aim to more fully sympathize—in a Smithian sense—with the world around us. Already many businesses are interested in technology that can perform affective computing, as this allows for machine learning vis-à-vis human sentiment—presumably the usefulness of understanding human emotion is apparent to those funding such affective computing projects. While it can be used for businesses to ensure more targeted advertising, it also has enormous potential for furthering research on autism and the like. This potential also raises questions in terms of how much work can we export to machines before it will affect the way our brains works. We may have calculators to add up sums for us, but we still see it as educationally essential to understand the dynamics behind addition. In an analogous way, it seems that what could be lost when the direct link between donor and cause is removed would be an opportunity to further empathize, learn, and discover for ourselves better ways of effecting meaningful change locally and globally.

REFERENCES

- Jenni, K., Loewenstein, G. 2007. "Explaining the 'Identifiable Victim Effect'" *Journal of Risk and Uncertainty* 14: 235–257.
- Kogut, T., Ritov, I. 2005. "The 'Identified Victim' Effect: An Identified Group, or Just a Single Individual?" *The Journal of Behavioral Decision Making* 18 (3): 157–167.
- Genevsky, A., Vastfjall, D., Slovic, P., and Knutson, B. 2013. "Neural Underpinnings of the Identifiable Victim Effect: Affect Shifts Preferences for Giving." *Journal of Neuroscience* 33 (43): 17188-17196.
- Forman-Barzilai, F. 2010. *Adam Smith and the Circles of Sympathy: Cosmopolitanism and Moral Theory*. Cambridge: Cambridge University Press.
- Small, D., Loewenstein, G., and Slovic, P. 2006. "Sympathy And Callousness: The Impact Of Deliberative Thought On Donations To Identifiable And Statistical Victims." *Organizational Behavior and Human Decision Processes* 102: 143–153.
- Smith, A. 1976a [1776]. *An Inquiry into the Nature and Causes of the Wealth of Nations*. Edited by R. H. Campbell, A. S. Skinner and W. B. Todd. Oxford: Oxford University Press.
- Smith, A. 1976b [1759]. *The Theory of Moral Sentiments*. Edited by D.D. Raphael and A. L. Macfie. Oxford: Oxford University Press.

Incommensurability and Partial Reference

Daniel P. Flavin

Hope College

ABSTRACT

The idea within the causal theory of reference that names hold (largely) the same reference over time seems to be invalid as concepts of scientific kinds have evolved. If science progresses by correcting mistakes of earlier scientists, then it must be possible to translate a different scientific theory's terms from either side of a scientific revolution. If the new theories and those they replace do not mean the same things by the terms they use, it appears as though we cannot straightforwardly say that the latter theory denies what the earlier theory asserts, in which case we cannot say that it represents a correction and improvement upon an earlier theory. Incommensurability is itself closely tied with the idea of translation and interpretation. Kuhn argues that causal theories of reference merely interpret instead of translate scientific terms, and thus scientific terms on either side of a scientific revolution remain incommensurable. If that is so, then this view holds profound consequences for scientific realism. This paper attempts to demonstrate how partial reference is a pragmatic way in which to interpret the different terms on either side of a scientific revolution thus saving at least one aspect of scientific realism from incommensurability.

KEYWORDS

Partial Reference, Incommensurability, Kuhn, Causal Theory of Reference, Translation, Interpretation

ACKNOWLEDGEMENTS

I would like to thank my family and friends for supporting my love of philosophy. I would also like to thank my professor Dr. Joseph LaPorte of Hope College for his help and guidance throughout the research and writing process of this paper and all those in attendance at the two conferences which this paper was presented at for their thoughtful comments and questions.

The idea within the causal theory of reference that names hold (largely) the same reference over time seems to be invalid as concepts of scientific kinds have evolved. If science progresses by correcting mistakes of earlier scientists, then it must be possible to translate a different scientific theory's terms from either side of a scientific revolution. If the new theories and those they replace do not mean the same things by the terms they use, it appears as though we cannot straightforwardly say that the latter theory denies what the earlier theory asserts, in which case we cannot say that it represents a correction and improvement upon an earlier theory. Thomas Kuhn holds this view that scientific terms are incommensurable, that they cannot be truly comprehended on either side of a scientific revolution. For Kuhn, incommensurability is itself closely tied with the idea of translation and interpretation. Kuhn argues that causal theories of reference merely interpret instead of translate scientific terms, and thus scientific terms on either side of a scientific revolution remain incommensurable. If that is so, then this view holds profound consequences for scientific realism.

In the first section of this paper I will discuss how incommensurability functions. In the second section of this paper I will discuss certain assumptions of incommensurability and scientific realism as well as demonstrate how partial reference offers a pragmatic way in which we can partially understand how past scientists attached scientific terms to reality thus saving at least one aspect of scientific realism though I argue that this understanding does not allow true comprehension of terms on either side of a scientific revolution.¹

I. SCIENTIFIC INCOMMENSURABILITY

Kuhn argues that the meanings of scientific terms and concepts often change within the theory in which they are deployed. Thus, it is impossible to define all the terms of one theory in the vocabulary of the other. To describe this phenomenon Kuhn borrowed the mathematical term 'incommensurability'. Applied to scientific terms incommensurability functions metaphorically; the phrase, "no common measure" becomes "no common language". The claim that two theories are incommensurable is then the claim that there is no language, neutral or otherwise, into which both theories, conceived as sets of sentences, can

1. I would like to thank my philosophy of science class at Hope College as well as the attendees of the Michigan Undergraduate Philosophy Conference and the Eastern Michigan University Undergraduate Philosophy Conference for their questions and comments which lead me to change my mind and subsequently my thesis.

be translated without residue or loss. Kuhn's incommensurability thesis therefore hinges on the difference between translating and interpreting terms (Kuhn 1982, 670).

Interpretation is not the same as translation. They are easily confused because actual translation often involves at least a small interpretive component. Translation is something done by a person who knows two languages. The translator systematically substitutes words or strings of words in one language for words or strings of words in such a way as to produce an equivalent utterance, whether it be vocal or written, in the other language. There are two important features of translation. First, the language into which the translation is made must exist before the translation began. That is, translation cannot change the meanings of the words or phrases examined. It may have increased the number of known referents of a given term, but it cannot alter the way in which those referents are determined. Second, that translation consists exclusively of words and phrases that replace, not necessarily one-for-one, words and phrases in the original. It is extremely important for Kuhn that translations must preserve not only reference but also sense or intension (Kuhn 1982, 681).

Interpretation only requires the knowledge of a single language. Interpretation is the process by which the use of a term is discovered. Once it has been completed and the word acquired, we use it and teach it to others. Translation does not occur. When an interpreter succeeds, what he or she has in the first instance done is learn a new language, one in which terms functioned differently. But acquiring a new language is not the same as translating from it into one's own.

For example imagine an American with no knowledge of any French terms or their referents living in Paris. This American is presented with a long loaf of bread with a hard crust and hears the term 'baguette'. Observing the circumstances around the utterance of the term 'baguette' he forms a hypothesis akin to 'baguette' means long loaf of bread with a hard crust. What that American has done in this instance is simply to learn a new language, one in which 'baguette' is a term.

That American may not merely introduce the term 'baguette' into her own language, English. That introduction would alter English and the result would not be a translation. Though English speakers may learn to use the term 'baguette', they speak the original language when they do so. When English speakers use

French terms such as 'esprit de corps' it is because the long English paraphrases for these French terms do not provide a true substitute since they are terms with inherent cultural meanings, meanings that can only be expressed within the cultural language which the term is employed. This cultural meaning is part of the sense and intension of the term 'esprit de corps' which would be lost in its translation into English. Thus, terms such as 'baguette' and 'esprit de corps' are incommensurable in the English language since translation is not achieved without a loss of both sense and intension.

To further illustrate how incommensurability affects a possible translation, like the example given above, Kuhn discusses the Newtonian term 'mass'. When learning Newtonian mechanics the terms 'mass' and 'force' must be acquired together, and Newton's Second Law must play a role in their acquisition. 'Mass' and 'force' cannot be learned independently and then used empirically to discover that force equals mass times acceleration. Nor can 'mass', or 'force', be first learnt and then used to define 'force', or 'mass' respectively with the aid of the Second Law. In formalizing mechanics one may select either 'mass' or 'force' as primitive and then introduce the other as a defined term. But that formalization supplies no information about how either the primitive or the defined terms attach to nature, how the forces and masses are picked out in the actual physical situations (Kuhn 1982, 677).

Though 'force' may be a primitive in some particular formalization of mechanics, one cannot learn to recognize forces without simultaneously learning to pick out masses and without recourse to the Second Law. That is why Newtonian 'force' and 'mass' are incommensurable with a physical theory, say Einsteinian, in which Newton's Second Law does not apply. Kuhn argues that 'mass' and 'force' simply cannot mean the same thing in both theories since Newton's Second Law is required for their definition in Newtonian mechanics but not in Einsteinian mechanics. Translation is impossible because no neutral language can be created. To learn either way of doing mechanics, Newtonian or Einsteinian, the interrelated terms, 'mass' and 'force', must be learned or relearned together and similarly applied, not in a piece meal fashion, but on nature whole.

Similarly, before Einstein's special theory of relativity, physicists accepted many assertions involving the term 'mass' that are not longer accepted today. The novelty of Einstein's theory is that Newton's assertions involving 'mass' were given

up. It was found that the referent of the Newtonian term 'mass' was indeterminate. It either referred to relativistic mass or proper mass. But, the physical referents of Einsteinian special relativity, proper mass or relativistic mass, are by no means identical with those of the Newtonian concept that bear the same name, 'mass'. Therefore Newton was not referring either to proper mass or relativistic mass, but instead he was referring to a non-existent quantity called "Newtonian mass" which has some properties of each. Therefore, Kuhn argues that after the scientific revolution brought about by Einsteinian special relativity, the Newtonian term 'mass' is incommensurable. However, partial reference, an idea within the causal theory of reference, offers a the scientific realist a pragmatic way in which the referent of 'mass', and other scientific terms in general, can be interpreted and understood.

II. A CAUSAL RESPONSE: PARTIAL REFERENCE

If scientific incommensurability exists then it would be impossible to say that theories evolve over time and that we come closer to the truth through that process, which is against the belief of scientific realism. Scientific realism can be described as such: the aim of science is to arrive at the truth about the world; scientific progress consists in progress toward that aim. The world which we inhabit, and which science investigates, is an objective reality, it exists independently of human cognitive activity. The result of successful scientific investigation is knowledge of both observable and unobservable aspects of the world. Scientists discover facts about unobservable entities whose behavior is responsible for the behavior of observable entities. Scientists propose theories which refer to unobservable entities in order to explain observed phenomena. As science progresses, theories approach the truth by providing increasingly accurate descriptions of entities identified by earlier scientists. Truth then is a relation of correspondence between language and reality. Whether a claim about the world is true is an objective matter. It depends on how things are in the mind-independent world, rather than on what scientists believe to be the case (Sanky 2009, 197).

In order to determine whether successive theories approach truth, the content of those theories must be compared, but comparison of the content of theories requires that the terms employed by those theories refer to the same object. In order for there to be an increase in truth known about a field of investigation,

successive theories must refer to a common domain of entities. Thus progress requires continuity of reference between theories. Scientific incommensurability argues that this continuity of reference does not hold and therefore the realist account of scientific progress as an increase in truth about a common domain of entities seems untenable. If later theories do not refer to the same entities to which earlier theories in the same domain referred, then it is not possible for later theories to increase the truth known about the same entities as those referred to by earlier theories. Under such circumstances, progress for a scientific realist is impossible; replacement of one theory by another is unable to constitute progress toward the truth about a common domain of entities.

However, for the scientific realist, semantic variance does not entail incomparability of content. Theories whose terms share reference may agree or disagree with respect to specific assertions even if the terms differ in sense. Co-reference of constituent terms is all that is needed for assertions to agree or disagree. Scientific theories may be compared with respect to content, provided only that the terms employed by the theories are related to a common domain.

Sanky argues that only in exceptional cases of wholesale ontological error is there any serious prospect of total incomparability of content. If theories are genuinely applied to the same domain, then, given the role of pragmatic factors in causal reference determination there will always be at least some overlap in the reference of the terms employed by theories in relation to the common domain such that they can be understood (Sanky 2009, 197).

In the causal theory of reference, reference is initially fixed with a dubbing, usually by perception. Reference is fixed during the initial baptism when a speaker says of a perceived object: "You are to be called 'N'." Reference is fixed via description when a speaker stipulates: "Whatever the unique such-and-such is to be called 'N'." After this initial reference fixing, the name is passed on from speaker to speaker through communicative exchanges. Speakers succeed in referring to something by means of its name because underlying their uses of the name are links in a causal chain stretching back to the dubbing of the object with that name. Subsequent speakers borrow their reference from speakers earlier in the chain. All that is required is that borrowers are appropriately linked to their lenders through causal chains of communication and that reference remains the same over time.

Partial reference theorists like Field, however, argue that many scientific terms are referentially indeterminate. It can be clearly seen throughout the history of science that words do change reference. For example, the term 'water', as it was once used, referenced the principle element water. The term 'water', as it is currently used, references the compound H_2O . Science is replete with cases such as this. That is because until 'water' was found out to be H_2O its referent was indeterminate. Many scientific terms are referentially indeterminate. If they are singular terms, there is no fact of the matter as to what they denote. If they are general terms there is no fact of the matter as to what their extension is. This indeterminacy exists until the scientific community negotiates and sets the term's referent. Newtonian 'mass' is one such example. After Newtonian mechanics was replaced by the special theory of relativity, Newtonian 'mass' could either denote proper mass or relativistic mass (Field 1973, 462).

After the advent of the special theory of relativity the Newtonian term 'mass' still refers, (i.e, maintains continuity of reference), albeit partially, to both relativistic mass and proper mass. Using the structure of the sentence, the term 'mass' can even be applied on a case by case basis to determine truth values. But this case by case determination of truth values raises many issues along with the fact that partial reference sometimes produces unsatisfactory answers.

Take for example the term 'mass'. When used in Newtonian mechanics 'mass' is the proportion constant between force and acceleration. Now with the advent of relativity theory 'mass' refers to either "proper mass" or "relativistic mass". Whether Newton was referring to proper mass or relativistic mass when he uses the term 'mass' can be determined using the sentence structure. So we then substitute the referentially indeterminate word, in this case 'mass', for its referentially determinate counterpart, which may lead to an unsatisfactory answer.

For example, to accelerate a body uniformly between any pair of different velocities, more force is required if the mass of the body is greater. In this example the term 'mass' partially refers to both relativistic mass and proper mass, no precise answer is achieved. The term 'mass' could not have referred both to proper mass and relativistic mass to Newton for obvious reasons, mainly Newton did not know that such things as proper mass and relativistic mass existed. Therefore, partial reference allows only for an ex post facto examination of past scientific theories with modern scientific knowledge in order to explain why some theory's hypotheses were confirmed by experience. Within a common domain of

scientific inquiry this pragmatically allows for the kind of progression towards truth which is the keystone of scientific realism.

Another problem arises with regards to partial reference mainly that examining the referential success or failure of scientific terms on an individual basis disregards how scientific terms are interrelated within any particular scientific theory. Think of the terms 'mass' and 'force' in Newtonian mechanics. They must be learned together and the theory within which they are deployed must be applied on nature whole, not in the piece meal fashion partial reference purports.

As stated before, Kuhn argues that translation must preserve not only reference but also sense or intension. In matching terms with their referents, a person may make use of anything a person knows or believes about those referents. Two people may speak the same language and nevertheless use different criteria in picking out the referents of its terms. This is because language is adapted to the social and natural world in which people live, and that world does not present the sorts of objects and situations which would, by exploiting their criterial differences, lead them to make different identifications. Members of the same language community are members of a common culture, and each member may therefore expect to be presented with the same range of objects and situations. If they are to co-refer, each must associate each individual term with a set of criteria sufficient to distinguish its referents from other sorts of objects or situations which the community's world actually presents (Kuhn 1982, 682).

It is essential that sets of terms must be learned together by those raised inside a culture. Foreigners encountering that culture must consider those terms together during interpretation, otherwise they are incommensurable. Inherent in this assumption is the fact that different languages impose different structures on the world. As Kuhn writes:

Imagine, for a moment, that for each individual a referring term is a node in a lexical network from which radiate labels for the criteria that he or she uses in identifying the referents of the nodal term. Those criteria will tie some terms together and distance them from others, thus building a multi-dimensional structure within the lexicon. That structure mirrors aspects of the structure of the world which the lexicon can be used to describe,

and it simultaneously limits the phenomena that can be described with the lexicon's aid....

Note, now, that homologous structures, structures mirroring the same world, may be fashioned using different sets of criterial linkages. What such homologous structures preserve, bare of criterial labels, is the taxonomic categories of the world and the similarity/difference relationships between them...What members of a language community share is a homology of lexical structure. Their criteria need not be the same, for those they can learn from each other as needed. But their taxonomic structures must match, for where structure is different, the world is different, language is private, and communication ceases until one party acquires the language of the other (Kuhn 1982, 682).

It is this emphasis on the social and cultural foundation of language which, I believe, makes the strong thesis of partial reference, i.e. that partial reference entirely defeats incommensurability as a translation schema, untenable. Partial reference allows the scientific realist to use modern scientific knowledge to interpret terms on either side of a scientific revolution but the understanding of those scientific terms can only be partial due to the fact we are removed from the lexical network within which the previous theory functioned. Scientific terms remain truly untranslatable, i.e. they lose sense and intension, but are made pragmatically understandable to our modern scientific knowledge base.

III. CONCLUSION

Kuhn argues that meanings of scientific terms and concepts often change with the theory in which they are deployed. Thus, it is impossible to define all the terms of one theory in the vocabulary of the other. To describe this phenomenon Kuhn borrowed the mathematical term 'incommensurability'. Applied to scientific terms 'incommensurability' functions metaphorically. The phrase, 'no common measure' becomes 'no common language'. For Kuhn the claim that two theories are incommensurable is then the claim that there is no language, neutral or

otherwise, into which both theories, conceived as sets of sentences, can be translated without residue or loss.

If scientific incommensurability exists then it would be impossible to say that theories evolve over time and that we come closer to the truth through that process, which is against the beliefs of scientific realism. However, partial reference offers a pragmatic way in which different scientific terms from theories on either side of a scientific revolution can be partially understood allowing for the progression necessary for scientific realism. In this way partial reference provides a realist response to explain how terms can be understood between different scientific theories as knowledge progresses.

REFERENCES

- Field, H., 1973. "Theory Change and the Indeterminacy of Reference." *Journal of Philosophy* 70 (14): 462–481
- Kuhn, T., 1982. "Commensurability, Comparability, Communicability." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Vol. 1982, Volume Two: Symposia and Invited Papers: 669-688.
- Sankey, H. 2009. "Scientific Realism and Philosophy of Science." *Studies in History and Philosophy of Science* 40: 196–202.

Seemingly Alive: Full-brain Death Is Death

Andrew Rydlund

University of St. Thomas, Minnesota

ABSTRACT

According to the Thomistic Hylomorphic approach to the human person, I argue that full-brain death results in the substantial corruption of the human being, or in other words, full-brain death sufficiently renders the body unfit to be informed by the rational soul. In like manner, I consider objections originating from philosophers who take on the Thomistic Hylomorphic approach to the human person, since among such hylomorphists there is disagreement as to whether full-brain death is sufficient for rendering the matter unfit for the soul or not. The argument I offer roughly goes as follows: according to Thomistic principles, if the body necessarily ceases to function as a unified whole with intrinsic finality after full-brain death occurs, then, according to Thomistic Hylomorphism, full-brain death sufficiently renders the matter unfit for the soul. I then show that the body does necessarily cease to function as a unified whole with intrinsic finality after full-brain death occurs by drawing from Dr. James Bernat's medical experience and research. Consequently, I conclude that according to Thomistic Hylomorphism, full-brain death sufficiently renders the matter unfit for the soul, or in other words, full-brain death results in the death of the human being. I then consider objections from and provide responses to two advocates of the Circulatory/Reparatory theory, who hold that as long as the circulation of oxygenated nutrient-rich blood continues, the matter retains its proper disposition for the soul and thereby constitutes that the human being is still present, even if the circulation is maintained via machines and pharmacotherapy. The soundness of my argumentation brings with it bioethical implications with regard to organ harvesting and donation, as well as a bearing upon when it becomes morally acceptable to terminate 'life support,' since, if any organized brain activity remains, even in the brain-stem alone, then there is still a human person present and alive.

KEYWORDS

Full-brain Death, Hylomorphism, Substance, Accident, Prime Matter, Substantial Form, Accidental Form

I. INTRODUCTION

In this paper I will argue that full-brain death results in the substantial corruption of the human being, or in other words, full-brain death renders the body unfit to be informed by the rational soul according to Thomistic Hylomorphic principles. Likewise, I will consider objections originating from philosophers who take on the Thomistic approach to the human person, since among such hylomorphists there is disagreement as to whether full-brain death is sufficient for rendering the matter unfit for the soul or not. I will argue as an advocate of the position which holds full-brain death as sufficient to render the matter unfit for the soul.¹

For St. Thomas Aquinas, the death of the human person occurs when the substantial form, the soul, is separated from the matter, namely, the body due to the matter becoming unfit to be informed by the soul. As stated by Jason Eberl, Aquinas holds the rational soul to be the “principle of the body’s (1) existence (esse), (2) unified organic functioning, and (3) specific nature as a ‘human’ body” (Eberl 2006, 44). Due to the first two ways in which the soul acts as principle of the body, Aquinas believes that to live can be taken in two distinct senses, saying, “In one, it means the very act of existence of that which is living, which rests on the union of the soul to the body as its form. In the other sense, ‘to live’ is taken to mean the activity of life” (Aquinas 13th Century, XIII.4.ad 2.). Accordingly, death can be taken in two senses as well, in one sense as the separation of the soul from the body and in the other as the cessation of the operations of life. Aquinas sees these two interpretations of death as transpiring coincidentally, since “the principle of a thing’s unity is the same as that of its being; for one[ness] is consequent upon being” (Aquinas 13th Century, II.58.5). Subsequently, when the soul ceases to inform the matter and thereby ceases to give it existence, the body will soon dissolve, since it no longer has an intrinsic unifying principle. Consequently, when “the body is no longer able to actualize [even] the soul’s vegetative capacities” (Eberl 2006, 45) death has occurred and no intrinsically directed functions remain

1. After presenting at the University of Michigan-Flint Philosophy Conference on Cognition and Neuroethics, I came to understand that I ought to explain better what I mean by “full-brain death.” I -do spell out exactly what I mean by full-brain death, however, for most this definition is easily confused with being in a vegetative state. Thus, I think it would be good for me to clarify right from the start that by “full-brain death,” I mean the death of the whole brain and this is distinct from a vegetative state because, one who is in a vegetative state still has a functioning brainstem, while one who is fully brain-dead does not.

between the organs. In summary, Aquinas teaches that death occurs when the matter, losing the active capacity to actualize even the soul's vegetative powers, becomes unfit to be informed by the soul which results in the body's loss of an intrinsic finality that is apt of a unified whole.

According to the abovementioned principles, I argue that full-brain death is indeed a sufficient criterion for rendering the matter unfit to be informed by the soul. The argument states that after full-brain death the body or corpse no longer functions as a unified whole. However, if the aforementioned is true, then full-brain death is sufficient for rendering the matter unfit to be informed by the soul. Thus, it follows via modus ponens that full-brain death is sufficient for rendering the matter unfit to be informed by the soul.

II. MAIN ARGUMENT

Now I will formally present the main argument of this paper and employ Aquinas' hylomorphic principles to support its premises. Whether or not these principles are true is irrelevant for the soundness of my argument, since I am simply arguing that according to the Thomistic Hylomorphic view of the person, full-brain death is a sufficient criterion for rendering the matter unfit for the soul.² The argument formally goes as follows.

1. If the body necessarily ceases to function as a unified whole with intrinsic finality after full-brain death occurs, then, according to Thomistic Hylomorphism, full-brain death sufficiently renders the matter unfit for the soul.
2. It is the case that the body necessarily ceases to function as a unified whole with intrinsic finality after full-brain death occurs.

2. It became clear at the conference that many people are not familiar with Thomistic thought and they thereby found my conclusion to be too narrow. A helpful suggestion was given to me with regard to my thesis in order to help it bear more weight, namely to mention that the Thomistic account of human persistence conditions does a better job than the other accounts of the human person (ie. Psychological theory, animalist theory...etc.), and if people wanted reasons as to why that is the case they could bring that up during the Q&A time.

3. Therefore, according to Thomistic Hylomorphism, full-brain death sufficiently renders the matter unfit for the soul.
(Modus Ponens 1&2)

In support of premise one I assert that if the body necessarily ceases to function as a unified whole with intrinsic finality, that is if it no longer has functions being actualized by an internal source of movement, after full-brain death occurs, then the soul has ceased to inform the body. Jason Eberl asserts, "When the union of a rational soul and its body is dissolved, the dissolution of the body's unified organic functioning immediately follows" (Eberl 2006, 44) since as Aquinas states, "If that which holds the individual contrary parts together is removed, they tend toward what is fitting to them according to nature, and thus the dissolution of the body is brought about" (Aquinas 13th Century, XXV.6 reply). Thus, when the body ceases to function as a unified whole with intrinsic finality, it follows that the soul no longer informs the body. Yet, the soul continues to inform the body as long as the matter is fit for the soul, since it is only "with the cessation of the requisite dispositions," namely the active capacities to carry out vegetative functions, that "a form cannot remain in the matter" (Aquinas 13th Century, XIV. ad 20). Thus, it follows via hypothetical syllogism that if full-brain death necessarily leaves the body without unity and intrinsic finality, then it is sufficient for rendering the matter unfit for the soul.

On behalf of premise two, James Bernat M.D., in defending whole-brain death as a sufficient criterion for death, demonstrates how all necessary critical functions for life are tied to the brain, saying, "Destruction of a critical array of neurons within the 'whole' brain...is necessary for death because: 1) the vital functions of respiration and control of circulation are subserved by the brain stem; 2) the critical integrative functions are subserved by the brain stem and hypothalamus; and 3) the wakefulness component of consciousness is subserved by the brain stem and the awareness component of consciousness by the thalamus and cerebral cortex" (Bernat 1998, 18). Thus, since all of the critical functions necessary for the life of an organism, including those of the vegetative type, are contingent upon the functioning of the brain, it follows that after full-brain death has occurred, which refers to "the complete cessation of all organized neurological activity throughout the entire brain, including the cerebrum, cerebellum, and brain stem," (NCBE 2014 FAQ) the circulation of blood, the respiratory system, and all other vitally integrative functions among the organs

necessarily cease to operate in a unified and intrinsically directed way. This necessary cessation of integrative function entails that after full-brain death the body no longer has the capacity to act as a unified whole with some intrinsic principle of finality. Thus, it follows from premises one and two via Modus Ponens that according to Thomistic Hylomorphism, full-brain death sufficiently renders the matter unfit for the soul.

III. OBJECTIONS AND RESPONSES

Here I will consider objections to my main argument and provide responses to them. The objections I will consider come from Dr. Mark Spencer and Alan Shewmon M.D. who advocate the circulatory/respiratory theory, which states that the matter remains fit for the soul as long as the circulation of oxygenated nutrient-rich blood continues. Both stand in opposition to the second premise of my main argument which asserts that the body ceases to function as a unified whole with intrinsic finality after full-brain death has occurred.³

The first objection I consider is from Dr. Spencer. He concludes from the circulatory/respiratory theory that the body does not necessarily cease to function as a unified whole with intrinsic finality after full-brain death occurs. I take his argument to go as follows:

1. The body continues to function as a unified whole with intrinsic finality as long as the circulation of oxygenated nutrient-rich blood continues.
2. If the body continues to function as a unified whole with intrinsic finality as long as the circulation of oxygenated nutrient-rich blood continues, then the body does not necessarily cease to function as a unified whole with intrinsic finality after full-brain death occurs.

3. Though there is a third view held among Hylomorphists asserting that higher-brain death by itself is a sufficient criterion for rendering the matter unfit for the soul, it does not have any bearing upon my argument, since if the higher-brain is sufficient, then the whole brain is sufficient as well.

compos mentis

3. Therefore, the body does not necessarily cease to function as a unified whole with intrinsic finality after full-brain death occurs. (Modus Ponens 1&2).

In support of premise one, Spencer makes reference to Aquinas' belief in a primary organ, asserting, "It is the organ that provides the motive power and the vital disposition necessary for all the organs to be alive by moving the necessary 'vital heat' and 'spirits,'" thus it appears that what is necessary for the matter to remain disposed for the soul is not the heart itself, but what the heart provides, namely, 'vital heat' and 'spirits', which today "can, due to their connection with breath, be equated in modern terms with oxygenated nutrient-rich blood" (Spencer 2010, 864). Furthermore, Spencer claims that due to technology allowing for circulation to continue after full-brain death has occurred, the body thereby retains its proper disposition for the soul and therefore is informed by the soul even after full-brain death occurs. Spencer concludes that "the fluid that [the heart] moves is thus a vegetative function," and accordingly, the body continues to function as a unified whole with intrinsic finality (Spencer 2010, 861).

Premise two is supported by the same principles that support the first premise of my main argument above (see section II Paragraph 2). If there is still a unified whole with intrinsic finality present, then it follows that the soul is still informing the matter. Yet, if that's the case after full-brain death has occurred, then the body does not necessarily cease to function as a unified whole with intrinsic finality after full-brain death occurs. Thus, it follows via Hypothetical Syllogism that if there is still a unified whole with intrinsic finality present after full-brain death, then the body does not necessarily cease to function as a unified whole with intrinsic finality after full-brain death occurs. Therefore, Spencer's conclusion follows via Modus Ponens from premises one and two.

In response to Spencer I reject premise one. Though I do concede that the circulation of oxygenated nutrient-rich blood is necessary for the continuation of life in most instances, I do not think it is sufficient for maintaining life after full-brain death occurs. I hold that after full-brain death occurs all intrinsically directed activity ceases, and thus, there is no longer a living organism present. The response goes as follows:

1. If the body continues to function as a unified whole with intrinsic finality as long as the circulation of oxygenated

nutrient-rich blood continues, then the organism will maintain activity that is intrinsically directed.

2. The organism does not maintain activity that is intrinsically directed.
3. Therefore, the body does not continue to function as a unified whole with intrinsic finality as long as the circulation of oxygenated nutrient-rich blood continues. (Modus Tollens From 1&2).

Dr. Spencer's claim itself stands in support of premise one:

Thus, some activity of the body must be able to be attributed to an internal source; there must be some activity of the whole organism that cannot be entirely explained in terms of external forces, such as the operations of a machine, if the organism is to be considered alive and still informed by the same soul it had when it was uncontroversially alive. (Spencer 2010, 866)

Thus, if the organism is still informed by the soul, as would be the case if it continues to function as a unified whole with intrinsic finality, then at least some activity or function of the organism is going to be intrinsically directed.

In defense of Premise two of my response, I argue that the apparent integrative activities such as "bodily growth, maintenance of homeostasis, digestion, excretion, and the fighting of diseases" (Spencer 2010, 866) do not result from the organism's functioning as a unified whole with intrinsic finality. What remains after full-brain death has occurred is completely dependent upon mechanical and pharmacological support, and thus, as claimed by Eberl, "The assumption of such functions by life-support machinery indicates that the body has lost the capacity to perform them under its own control" (Eberl 2006, 57). Thus, the body does not have continued metabolic function attributable to an internal source. Furthermore, since the machine regulates the actions of the organs, "such functions can be understood as emerging from the interaction of a body's organ systems without entailing that the body has the integrative unity required for it to compose an individual substance that is *unum simpliciter* with a

single substantial form" (Eberl 2006, 58). Thus, it follows from premises one and two via Modus Tollens that the body does not continue to function as a unified whole with intrinsic finality as long as the circulation of oxygenated nutrient-rich blood continues.

The second objection comes from Shewmon, who offers a counter example to the assertion that the body necessarily ceases to function as a unified whole with intrinsic finality after full-brain death occurs. Shewmon founds his example upon the assumption that the brain only acts as modulator of the integrative somatic functions in the body, and consequently is not necessary for them if artificial life support and pharmacotherapy are provided. I take his counter example to go as follows:

1. If a nutritive environment within the uterus of a fully brain-dead mother can be maintained until the fetus comes to term, then the body does not necessarily cease to function as a unified whole with intrinsic finality after full-brain death occurs.
2. A nutritive environment within the uterus of a fully brain-dead mother can be maintained until the fetus comes to term.
3. Therefore, the body does not necessarily cease to function as a unified whole with intrinsic finality after full-brain death occurs. (Modus Ponens 1&2)

In defense of his first premise, Shewmon responds to Corrado Manni who claims that such a woman acts only as a "biological incubator;" asserting that his description "does injustice to the complex, teleological, organism-level, physiological changes of pregnancy (weight gain, internal redistribution of blood flow favoring the uterus, immunologic tolerance toward the fetus, etc.), which occur despite the absence of brain function" (Eberl 2006, 55). Accordingly, Shewmon concludes that such processes could never occur if the organism were not alive and acting as a unified whole with intrinsic finality. Thus, if a fully brain-dead mother can bring a fetus to term, then the body does not necessarily cease to function as a unified whole with intrinsic finality after full-brain death occurs.

The case of Robyn Benson, of Victoria, British Columbia, proves premise two to be true. This woman suffered a brain hemorrhage and became fully brain-dead while she was 22 weeks pregnant, and the doctors kept her body 'alive' for 6 weeks until the baby was born premature, but still healthy. Thus, Shewmon's conclusion would seem to follow from premises one and two via Modus Ponens.

In response to Shewmon's objection I refute premise one, since the conditional statement on which his argument rests is flawed. That which maintains the nutritive uterus is not actually the mother, but the doctors who are using machines and medicine to do the work. My response goes as follows:

1. If the conditional in premise one of Shewmon's objection is true, then the mother's body should have at least some intrinsically directed metabolic function that is not caused by mechanical and pharmacological intervention.
2. The mother's body does not have at least some intrinsically directed metabolic function that is not caused by mechanical and pharmacological intervention.
3. Therefore, the conditional in premise one of Shewmon's objection is false. (Modus Tollens 1&2).

For support of premise one of my response see Section One Paragraph two, which states that a living thing has a soul which brings about some intrinsically directed functions.

In support of premise two, Eberl reveals that the fully brain-dead mother's body requires extensive work in order to maintain a properly nutritive uterus. Sighting Dr. Field, Eberl shows that the body of a 27-year-old fully brain-dead mother required "mechanical ventilation, vasopressors to treat fluid-resistant hypotension, warming or cooling blankets to treat temperature lability, nutritional support, replacement hormones to treat endocrine abnormalities, aggressive surveillance for and treatment of infections, and heparin prophylaxis," and furthermore, "maximum effort was directed at treating the severe hypotension, temperature fluctuations, diabetes insipidus, hypothyroidism, and cortisol deficiency that were thought to be the result of the autoregulatory function of the brain" (Eberl 2006, 57). Appropriately, the aforementioned evidence for the

extensiveness of the treatment necessary to provide a nutritive uterus makes it difficult to see how any intrinsically directed metabolic functions are being carried out. Rather, I assert that after full-brain death occurs, such functions are the result of mechanical and pharmacological influence that extrinsically impose finality upon the organs so that they work together to carry out metabolic functions, just as the parts of a watch work together to tell time. Thus, from premises one and two the falsehood of Shewmon's first premise follows via Modus Tollens, since the antecedent of his conditional can be true without the consequent being true.

IV. CONCLUSION

In conclusion, recall that I have presented Aquinas' approach to human death, which holds the soul to be the principle both of existence and of mover, and accordingly, it follows that when the soul separates from the body, the matter ceases to be intrinsically directed toward the actualization of its metabolic or vegetative functions. Furthermore, once the matter stops having the active capacity to actualize even the vegetative powers, the matter has become unfit for the soul and death occurs. Consequently, due to its effects, I concluded that full-brain death is a sufficient criterion for rendering the matter unfit for the soul. Next I considered objections from both Dr. Mark Spencer and Alan Shewmon who, advocating the circulatory/respiratory theory, attempt to disprove that the body necessarily loses its functional unity and intrinsic finality after full-brain death occurs. Yet in response to these arguments I demonstrated how fully brain-dead patients are not actually carrying out vegetative metabolic functions, but rather such functions are the result of the organ systems interacting with one another due to the influence of external machinery and medicine. Thus, I conclude that after full-brain death occurs, there is no longer any intrinsically directed functions being carried out by the body, but only functions under extrinsically imposed finality, and subsequently, full-brain death is a sufficient criterion for rendering the matter unfit for the soul.

REFERENCES

Aquinas, Thomas. 13th Century. "Thomas Aquinas: Quaestiones Disputatae De Veritate: English," *Thomas Aquinas: Quaestiones Disputatae De Veritate: English*. N.p., n.d. <http://dhsprory.org/thomas/QDdeVer.htm>.

- Aquinas, Thomas. 13th Century. "Quaestiones Disputate De Anima." *Thomas Aquinas: Quaestiones Disputate De Anima: English*. N.p., n.d. <http://dhspriority.org/thomas/english/QDdeAnima.htm>.
- Aquinas, Thomas. 13th Century "Contra Gentiles." *Dominican House of Studies Priory*. N.p., n.d. <http://dhspriority.org/thomas/ContraGentiles.htm>.
- Eberl, Jason T. 2006. *Thomistic Principles and Bioethics*. London, UK: Routledge.
- Bernat, James L. 1998. "A Defense of the Whole-Brain Concept of Death." *The Hastings Center Report* 28.2: 14-23. JSTOR.
- "Resources." *FAQ*. National Catholic Bioethics Center, n.d. <http://www.ncbcenter.org/page.aspx?pid=1285>.
- Spencer, Mark K. "A Reexamination Of The Hylomorphic Theory Of Death." *The Review of Metaphysics* 63.4 (2010): 843-870.

compos mentis

Identity and Causality: Foucault's Subject and Kant's Third Antinomy

Rebecca Valeriano-Flores

DePaul University

ABSTRACT

Foucault's work is concerned with the formation of the subject through self-knowledge: the individual is subject to dominating forms of power or the subject themselves. Pastoral power is a form of power that provides self-knowledge to the subject as identity while also convincing the subject that identity is essential to the individual and their body. To liberate the individual from these forms of power, the individual must create new forms of identity by developing their own self-knowledge. This is Foucault's ethics of freedom, which is similar to Kant's morality as it is related to the third antinomy. For Kant, causes through freedom and determined causal nature are compatible with each other. The individual can lead a moral, free life if morality is pulled out of the causal realm. By avoiding making choices as means to an end, thus pulling it out of the causal realm, one can make free and moral choices for the thing in itself. For Foucault, pastoral power belongs to the causal chain. To free the individual from the causal chain, one must practice ethics as the care of self, in which they form their own self-knowledge that avoids the essentialism of causal identity. Freedom is the necessary condition for these ethics. By liberating oneself from the chain of causality, the individual can create identity.

KEYWORDS

Causality, Identity, Ethics, Free Will, Determinism

INTRODUCTION

Foucault discusses Kant's "Answering the Question: What is Enlightenment?" essay¹ as a moment that asks about the present state of philosophy and modernity in general. He claims that this essay, for the first time, looks back on and examines ourselves. Foucault ([1982] 2000, 335) says, "[c]ompare this with the Cartesian question: Who am I? ... But Kant asks something else: What are we?" Foucault takes up the Kantian project that seeks to prevent philosophy from going beyond possible experience to make metaphysical claims, including the metaphysical claim of the essence of identity. For Foucault, this Kantian project is a critical project that looks back at our own subjectivity; he takes it a step further by returning to an analysis of empirical conditions in his genealogical work. Though Foucault isn't a transcendental philosopher, it can be considered that he continues Kant's critical project.

In section I, this paper will discuss Michel Foucault's work as it is primarily concerned with the formation of the subject. "Pastoral power" is a form of power that offers the subject self-knowledge; in believing the knowledge of themselves as offered by different disciplines and structures, the subject reproduces this knowledge in their action. Pastoral power also seeks to convince the subject that identity is essential to the individual. Foucault shows through his work that identity is not essential to the individual, but is created by the individual.

In section II, I will show that Foucault's ethics of freedom shares an important similarity with Kantian morality: they both believe that to live an ethical life, morality must be removed from the determinism of causal nature. Kant's morality is built on the third antinomy: are there causes through freedom (free will), or is there no freedom in causal nature (determinism)? Kant's answer is that both are compatible with each other, if appearances are attributed to phenomena and the thing in itself to noumena. He carries this through his morality: to live a moral life, the individual must extract morality from causal nature and it must be performed for the thing in itself. Foucault's ethics can be considered in a similar way: he believes that ethics itself is a practice of freedom. If pastoral power is part of causal nature, then the individual must liberate themselves from it through a practice of freedom.

1. Kant and the Enlightenment is discussed in Foucault's "What is Enlightenment?" (2000) and also "The Subject and Power" ([1982] 2000).

I. FOUCAULT: PASTORAL POWER AND ESSENTIAL IDENTITY

Michel Foucault's work is largely concerned with the formation of the subject. He explores this through two forms of subjectivity: the individual is subject to what he calls "pastoral power" of dominating institutions and disciplines, and simultaneously, the individual is subject to their own self-knowledge. Pastoral power, through its complex and totalizing methods, seeks to convince the subject that identity is essential to the individual.

A. Pastoral Power in its Early and Modern Forms

In Foucault's "The Subject and Power" essay and *Wrong-Doing, Truth-Telling* lectures at the Catholic University of Louvain, the form of pastoral power originated in the early Christianity of roughly the first through fifth centuries, A.D. Two of the main distinguishing aims of pastoral power of early Christianity seek to assure salvation in the afterlife, and cultivate the inner life or soul of the individual.² These aims were used as techniques of power that seek to convince the subject of a truth being produced. The church preached this: to be saved, the individual must atone for their sins through penance. This also concerns the entirety of the subject, viz. their life within the church, outside of the church, and even *after* their life on Earth. The individual, believing this truth, makes themselves a subject to the church by reproducing and confirming this truth, especially in the act of penance.

The aims of pastoral power needed to change so that it could reach beyond the church. Rather than assuring salvation in the afterlife, the aim was to assure salvation in the present life. Salvation took on different forms for different disciplines: for example, with the medical discipline, salvation is health; with the police state, salvation is security and safety of the population.

In *Wrong-Doing, Truth-Telling*, Foucault (2014) makes this connection by relating sin and penance (of Christian pastoral power) to disease and medicine (of modern pastoral power). The promise of salvation in the next life mirrors the promise of well-being for the patient; Foucault (2014, 183) says, "'penance was a medicine ... [o]ne was sick, and sin was an illness or sin was a wound.'" In *The Birth of the Clinic* ([1963] 1994) the gaze and verbalization of symptoms between the

2. Foucault also discusses two additional characteristics, the other two being: the form is prepared to sacrifice itself, and that it is concerned with both the community and the individual, for their entire lives (Foucault [1982] 2000, 333).

doctor and patient simultaneously gathered symptoms that point to and represent disease, while at the same time created the organization of the disease and its symptoms. A type of self-knowledge is offered to the patient, who is both subject and object of this knowledge. Through self-governing, the individual's belief of authority manifests in their actions: the individual performs an action, following the doctor's orders. The medical discipline then offers the individual knowledge of themselves, and the individual believes this knowledge and acts on it.

B. Self-Knowledge and Creating Identity

This knowledge also includes the notion that identity is essential to the individual. *The Birth of the Clinic* touches on the theme of essences. The medical discipline after the eighteenth century created a rational order of diseases with carefully and systematically organized symptoms. "It is a space in which analogies define essences," to the extent that "the order of disease is simply a 'carbon copy' of the world of life" (Foucault, [1963] 1994, 6-7). The medical discipline's ordered inventory of symptoms and diseases sets out to be an exact description of something that is extended within the body. The effect of this methodology is not only individualizing, but totalizing, viz., it applies to the individual patient and also all individual bodies.

Both pastoral power and Foucault's definition of the "subject" are clarified in his "The Subject and Power" essay ([1982] 2000, 331):

This form of power applies itself to immediate everyday life which categorizes the individual, marks him by his own individuality, attaches him to his own identity, imposes a law of truth on him which he must recognize and which others have to recognize in him. It is a form of power which makes individuals subjects. There are two meanings of the word "subject": subject to someone else by control and dependence; and tied to his own identity by a conscience or self-knowledge.

The disciplines that Foucault talks about throughout his work exemplify this form of power: the prison system, the education system, the family, the medical discipline, and the church use dominating methods not by explicitly coercing the subject, but by making the subject believe a truth of themselves as well as contributing to the individual's self-knowledge. Once the subject believes this

truth, they also reproduce it in their self-governed actions. The primary question for the subject is “Who am I?” or “Who are we?” and pastoral power offers an answer. The medical discipline’s answer is that the individual is essentially its anatomical systems, and any sickness must be expelled by a doctor in order for one to be well. For Christianity, the individual’s essence is their soul, and sin must be purged from it through penance in order to save it from the damnation of hell. Pastoral power is thus concerned with the self-knowledge of individual’s essence – in other words, the identity that is essential to the individual.

“The Subject and Power” and other essays and interviews also touch on intersecting notions of identity including gender and sexuality, and other philosophers and social theorists use Foucault’s tools of analysis in critical theory on race and class. The identity that these disciplines seek to dominate is not only the essence of the soul, but also the intersecting notions of identity at the center of current struggles. Individuals are stigmatized and punished based on identity that are managed by forms of pastoral power; identity is reproduced in action by the individual. Pastoral power does not rest in a singular entity like the state; rather, it is complex, multiple, and reproduced by each individual as they believe different truths of their identity.

But for Foucault, identity is not essential to the individual or to bodies. Through new forms of subjectivity, the individual can reject the identity that is the subject of forms of pastoral power. To return to “The Subject and Power,” Foucault says, “[m]aybe the target nowadays is not to discover what we are but to refuse what we are” ([1982] 2000, 336). He goes on:

The conclusion would be that the political, ethical, social, philosophical problem of our days is not to try to liberate the individual from the state, and from the state’s institutions, but to liberate us both from the state and from the type of individualization linked to the state. We have to promote new forms of subjectivity through the refusal of this kind of individuality which has been imposed on us for several centuries.

A new form of subjectivity would mean that the individual would not accept the identity that has been imposed on them by pastoral power, but instead create identity in action and becoming. Rather than the expose an essential truth of identity that is found within the self, new forms of subjectivity would emphasize

creativity, and would create culture and identity without the limits imposed by dominating forces.

II. FOUCAULT AND KANT

If the guiding question at hand is "Can we *freely* change our self-identity?" then Foucault's answer is yes, through these new forms of subjectivity, which Foucault later develops into his ethics. But to illuminate Foucault's ethics as they are related to the concept of freedom, they must be put into the light of Kant's ethics, specifically as they are related to the third antinomy and morality. For Kant, free will and determinism are compatible with each other when split into appearances of phenomena and the thing in itself of noumena. His morality depends on this as well: moral choices need to be pulled out of the causal nature of phenomena, and performed noumenally for the thing in itself. Foucault's ethics can be considered in a similar way. If pastoral power belongs to causal necessity, Foucault's ethics require the individual to liberate themselves from the causal chain. For Foucault, freedom is the necessary condition for an ethical life.

A. Kant's Third Antinomy and Morality

Kant's discussion of the third antinomy and its relevance to morality can be used to illuminate Foucault's ethics. In the *Prolegomena to Any Future Metaphysics* ([1783] 2001, 75) the third antinomy is this:

Thesis

There are in the world causes through freedom.

Antithesis

There is no freedom, but all is nature.

On one hand, the thesis states that there are causes through freedom, or to put it another way, there are causes that are not themselves an effect of another cause. On the other hand, the antithesis states that all is causal nature, in other words, every cause is itself an effect of a preceding cause, and that preceding cause is itself an effect, ad infinitum. For Kant, both the thesis and antithesis are compatible with each other, if causes through freedom is attributed to things in

themselves (noumena) and causal nature is attributed to the appearance of things (phenomena).

Now I may say without contradiction that all the actions of rational beings, so far as they are appearances (encountered in some experience), are subject to the necessity of nature; but the same actions, as regards merely the rational subject and its faculty of acting according to mere reason, are free (Kant [1783] 2001, 80).

This follows through to his morality in *Grounding for the Metaphysics of Morals* ([1785] 1993). The thesis suggests that we have free will; the antithesis suggests everything is determined. For Kant, this division of freedom and causality, or noumena and phenomena, is important to lead a moral life. Morality must be pulled out of phenomena and attributed entirely to noumena; it must be completely removed from the causality of the phenomenal realm.

In the *Grounding*, morality must follow this *a priori* formal principle rather than an *a posteriori* empirical incentive. In willing an incentive, there is no freedom, because "in willing an object as my effect there is already thought the causality of myself as an acting cause" ([1785] 1993, 27). For Kant, a maxim must be performed for the thing in itself, instead of a possible outcome of the maxim. In willing an effect, the person who does the willing already considers themselves as the cause; themselves as cause means they are also an effect of another cause. The individual willing an effect is then determined, and puts themselves into the chain of causality. But if the thing is done for itself, then the effect doesn't make a difference, and the individual is not in that chain. Furthermore, Kant goes on to develop this into the categorical imperative: only in the noumenal realm of the thing in itself can there be universal, objective moral law.

B. Foucault on Pastoral Power and Causality

Foucault takes a similar stance in that ethics must be removed from causality. Freedom, as expressed in the thesis of the third antinomy, is the necessary condition for leading an ethical life. Pastoral power belongs to the causal nature of the antithesis in the third antinomy.

1. Causality and Salvation

The causality of pastoral power is in its aim, the incentive of salvation. With the pastoral power of early Christianity, the purpose of penance is intended for the aim of salvation in the afterlife. By assuring the individual that sin is the truth of themselves, penance in Christianity works not only as a verbal admission that the individual is essentially sinful, but also the necessary act in order to save themselves and secure salvation.

[W]hat is important within Christianity seems to me to be far more this relationship to the individual to his truth than the problem of sin. ... What seems to me to have been the essence of Christianity and to have made a break in the history of Western subjectivity is the technique, the relationship of truth, and all the techniques put forward and perfected to draw out the truth of oneself with regard to sin (Foucault 2014, 117).

The individual admits that they are sinful in penance. In the *Wrong-Doing, Truth-Telling* lectures, Foucault describes the different forms of penance in early Christianity: from penance as a ritual performed only once during a person's life, to a life of penance as complete submission, and then finally to the form of confession that we're familiar with today. For the church to convince the individual that it can assure salvation, this admittance of sin by the individual is necessary. The individual must believe this truth --they are guilty of sin and that sin is within them. The affirmation of truth also affirms that penance is necessary to redeem them.

Pastoral power offers the individual a path to salvation in the afterlife; salvation is the willed incentive that Kant discusses in *Grounding*. The church's ethics are within the causal chain: in willing the *effect* of incentive of salvation, the individual in the act of penance is the *cause*. Though the church may offer a multitude of other ethical values, if according to Foucault one of the main distinguishing characteristics of pastoral power is the promise of salvation, then it lacks the moral content that Kant discusses in the *Grounding*. The act of penance or the confession of sin is not performed for the act in itself, or even the immediate effects of the act, but for an effect or incentive. Even if the afterlife is considered something outside possible experience, the aim of salvation in the afterlife concerns something essentially a part of the individual: their soul. This

also applies to the modern form of pastoral power. In the example of the medical discipline, salvation has expanded to mean health and well-being. The incentive or intended *effect* of health and well-being necessarily includes the individual and prescribed actions as a *cause*, therefore this is within the causality of the phenomenal realm. When individuals are made subjects to the modern pastoral power of the medical discipline, they are a part of the deterministic causal chain.

2. Causality and Identity

Causality in pastoral power can also be seen through Foucault's take on the semiotic system of signifier and signified. In *Birth of Clinic*, for the individual under the doctor's examination, the signifier was a symptom that not only pointed to the signified disease, but the order of the symptoms also completely represented the disease itself. This system is also applied to modern forensic psychiatry, which Foucault discusses in the sixth lecture of *Wrong-Doing, Truth-Telling*. The discipline of psychiatry extended to the criminal justice system with an eye on the criminal. It was no longer enough to consider the crime itself punishable; forensic psychiatry in the judicial system needed to ask *who* the criminal was. The crime itself would be a signifier of something signified within the criminal, like the symptom that signified a disease within the patient. The signified diseases were the psychiatric categories that developed at the end of the nineteenth century: necrophilia, kleptomania, exhibitionism, and homosexuality, among others. Foucault (2014, 221) goes on to describe this in terms of causality in psychiatry:

Madness then appeared to be the cause of that which made no sense, and the lack of responsibility established itself within that gap. But with this new analysis of instinct and emotions, there arose the possibility of a causal analysis of all conduct, whether criminal or noncriminal and whatever its degree of criminality. At this point, the juridical and psychiatric problem of crime entered an infinite labyrinth: if an act was determined by a causal nexus that the analysis of the criminal subject could uncover – if, then, an act was determined by such a causal nexus, could it be considered to be free?

The causality of psychiatric categories is explicit in the signifier-signified form, but perhaps less explicit in the contemporary terms that we are familiar with today. Homosexuality, for example, is no longer in the Diagnostic and Statistical Manual

of Mental Disorders, but remains as an identity of sexuality. A wide amount of research has concluded that homosexuality is the effect of biological cause, or possibly a mix of genetic, environmental, and hormonal factors (Frankowski 2004). Considered this way, identity belongs to causal nature, whether it's sexual identity, criminal identity, or cultural identity. In this form of psychiatric power and the wider form of pastoral power, identity is the effect of a cause. The causes range from biological to historical, from childhood trauma to diet, all depending on the form of power. Thus, identity as a result of these forms of power is entirely deterministic, and completely a part of the causal chain, the antithesis of the third antinomy.

Returning to the question "Can we *freely* change our self-identity?", if the thesis and antithesis of the third antinomy aren't compatible with each other, the answer would be *no*. If free will and determinism are not compatible, then our identities are determined by factors out of our control. But for Kant, freedom and causal nature are compatible, and one can lead a free and ethical life by withdrawing moral choices from causality and making these choices based on the thing in itself. Foucault says something similar in his own ethics of freedom.

C. Foucault on Ethics as a Practice of Freedom

Q. You say that freedom must be practiced ethically...

M.F. Yes, for what is ethics, if not the practice of freedom, the conscious [*réfléchi*e] practice of freedom?

Q. *In other words, you understand freedom as a reality that is already ethical in itself.*

M.F. Freedom is the ontological condition of ethics. But ethics is the considered form that freedom takes when it is informed by reflection (Foucault [1984] 1997a, 284).

For Foucault, freedom is the necessary condition of an ethical life. The issue of liberation from individualizing and totalizing forms of pastoral power should not aim to liberate people from an entity such as the state, but liberation should be manifested by the individual as a practice of freedom. Pastoral power imposes on the subject knowledge about themselves; to be free of the determinism of

pastoral power, the practice of freedom means the individual must make themselves a subject of their own self-knowledge and create new forms of identity. Rather than being limited to the identity produced by pastoral power's causal chain, one must form identity in new, multiple, and creative ways. Rather than believing that one must discover an identity that has been lurking within them, one should instead re-imagine and reconstruct what identity should mean for them. Foucault ([1984] 1997b, 163) expresses this in regards to sexuality:

Sexuality is a part of our behavior. It's a part of our world freedom. Sexuality is something that we ourselves create—it is our own creation, and much more than the discovery of a secret side of our desire. We have to understand that with our desires, through our desires, go new forms of relationships, new forms of love, new forms of creation. Sex is not a fatality: it's a possibility for creative life.

Sexuality and any form of identity—gender, race, and culture—is something that we create, not something that is essential to ourselves and the world. Furthermore, the development of self-knowledge and self-reflection that Foucault discusses in his ethics is influenced by the Stoic philosopher Seneca. He contrasts the different techniques that Stoic philosophy prescribes to form the subject with the system of penance and confession in early Christianity.

Self-examination for Seneca was not a search within the individual that resembles the search for an illness in the pastoral power of the medical discipline. The technique of self-writing is not like an intimate diary, narrative of life, or secret confession; it is an administrative, non-judgmental inventory. Instead of committing sin, the individual analyzes mistakes against their own ethical code, and makes a plan on how to avoid these mistakes in future action. Most importantly and in line with Foucault's ethics of freedom, these Stoic techniques also emphasize liberation and self-mastery. In *Wrong-Doing, Truth-Telling*, Foucault talks about about self-writing as an inventory of dependencies as seen in Seneca's letters to Seneca.

[This] is not at all an index of the presence in him of a hidden desire or concupiscence. It is simply an indicator of liberty. It is an indicator of liberty that allows him to say: "This is what I can do

without, and this is what I am still unable to give up” (Foucault 2014, 102).

This inventory of dependencies enables the individual to examine their liberation and also the things they still need to be liberated from. This is a part of a wider discussion on the care of self in Greek philosophy. Foucault describes the care of self as a practice of freedom as ethics. Not only does the individual need to “know thyself” but also to develop themselves independently. The care of self also requires the individual to take into consideration the different forms and methods of subjectivity, including those of dominating structures. By refuting the essentialism of pastoral power, the individual creates their own self-knowledge through self-reflection, and through action, agency, and becoming, creates new identity.

CONCLUSION

To conclude, Foucault’s work is concerned with the formation of the subject. Dominating forms of power impose self-knowledge onto the subject, which the subject must *believe* in order to reify. Pastoral power is a form of power that provides self-knowledge to the subject as identity while also convincing the subject that identity is essential to the individual and their body. To liberate the individual from these forms of power, the individual must create new forms of identity through their own self-knowledge. This is Foucault’s ethics of freedom, which is similar to Kant’s morality as it is related to the third antinomy. For Kant, causes through freedom and determined causal nature are compatible with each other. The individual can lead a moral, free life if morality is pulled out of the causal realm. By avoiding making choices as means to an end, thus pulling it out of the causal realm, one can make free and moral choices for the thing in itself. For Foucault, pastoral power belongs to the causal chain. To free the individual from the causal chain, one must practice ethics as the care of self, in which they form their own self-knowledge that avoids the essentialism of identity. Freedom is the necessary condition for these ethics. By liberating oneself from the chain of causality, the individual can create identity rather than become a subject to dominating forms of power.

The antinomy of freedom versus causal nature was an important concept in modern philosophical thought. While many later philosophers in the nineteenth century and beyond worked to supersede this duality, Kant’s critiques still

influence contemporary thought, even if the concepts have been broken apart and only tiny pieces remain, all scattered across theory. Foucault's work may contain some of these pieces, but cannot be considered transcendental philosophy, since many differences and rifts between the two philosophers remain, not to mention Foucault criticized transcendental philosophy in his own work.³

Future research on this subject will explore the similarities and differences between Kant and Foucault and also address current criticism. Contemporary critiques of Foucault claim that his ethics of freedom and creativity champions neoliberalism.⁴ But there's a difference between the ideology of individualism that we're familiar with through contemporary Marxist interpretations and the individual empowerment of Foucault's ethics. The ideology of individualism itself is a type of self-knowledge that the dominating form of power tries to impose on the subject. There are also the ideologies that perpetuate totalizing and individualizing racism, sexism, and homophobia. But societal change must start with the individual, with the self as subject. Foucault's ethics of freedom can be seen as a method of praxis, and can be compatible and useful for the politics of the radical left, especially when it comes to transforming thought into action. In the the preface to *Anti-Oedipus*, Foucault ([1976] 2000, 109) gives a guidelines for an anti-fascist handbook:

Do not think that one has to be sad in order to be militant, even though the thing one is fighting is abominable. It is the connection of desire to reality (and not its retreat into the forms of representation) that possesses revolutionary force.

Despite the criticism, Foucault's work stands as very compelling and influential among philosophers and theorists across disciplines. But nothing is safe from criticism, not even Foucault; and like Kant, his work has provided excellent tools to make those very critiques.

3. For a critique of Foucault as transcendentalist, see Colin Koopman's "Historical Critique or Transcendental Critique in Foucault: Two Kantian Lineages" (2010).

4. See recent criticism by Daniel Zamora in Jacobin (2014) and a discussion on Stuart Elden's blog, *Progressive Geographies* (2014).

POSTSCRIPT

This paper was submitted as a working draft and then later selected and read at the Michigan Undergraduate Philosophy Conference in 2015. I'd like to thank Thomas Mann and all of the participants at the conference for their excellent questions and comments. No major changes have been made since the original draft was submitted, except for a shortening of length to meet submission requirements.

REFERENCES

- Elden, Stuart. 2014. "Foucault and Neoliberalism – a few thoughts in response to the Zamora piece in Jacobin." In *Progressive Geographies*, December 17. <http://progressivegeographies.com/2014/12/17/foucault-and-neoliberalism-a-few-thoughts-in-response-to-the-zamora-piece-in-jacobin>.
- Foucault, Michel. (1982) 2000. "The Subject and Power." In *The Essential Works of Michel Foucault: Power*, edited by Paul Rabinow, 326-348. New York: The New Press.
- Foucault, Michel. 2000. "What is Enlightenment?" In *The Essential Works of Michel Foucault: Power*, edited by Paul Rabinow, 303-320. New York: The New Press.
- Foucault, Michel. (1976) 2000. "Preface to Anti-Oedipus." In *The Essential Works of Michel Foucault: Power*, edited by Paul Rabinow, 106-110. New York: The New Press.
- Foucault, Michel. (1984) 1997a. "The Ethics of the Concern for Self as a Practice of Freedom." In *The Essential Works of Michel Foucault: Ethics, Subjectivity, and Truth*, edited by Paul Rabinow, 281-302. New York: The New Press.
- Foucault, Michel. (1984) 1997b. "Sex, Power, and the Politics of Identity." In *The Essential Works of Michel Foucault: Ethics, Subjectivity, and Truth*, edited by Paul Rabinow, 163-174. New York: The New Press.
- Foucault, Michel. 2014. *Wrong-Doing, Truth-Telling: The Function of Avowal in Justice*, edited by Fabienne Brion and Bernard E. Harcourt, translated by Stephen W. Sawyer. Chicago: University of Chicago Press.
- Foucault, Michel. (1963) 1994. *The Birth of the Clinic: An Archaeology of Medical*

Perception, translated by A. M. Sheridan Smith. New York: Vintage Books.

Foucault, Michel. (1984) 1988. *The History of Sexuality, Vol. 3: The Care of the Self*, translated by Robert Hurley. New York: Vintage Books.

Foucault, Michel. (1976) 1980. *The History of Sexuality, Vol. 1: An Introduction*, translated by Robert Hurley. New York: Vintage Books.

Frankowski, Barbara L. and Committee on Adolescence. 2014. *Pediatrics* 113:1827-1832.

Kant, Immanuel. (1783) 2001. *Prolegomena to Any Future Metaphysics*, Second Edition, translated by James W. Ellington. Indianapolis/Cambridge: Hackett Publishing Company.

Kant, Immanuel. (1785) 1993. *Grounding for the Metaphysics of Morals*, translated by James W. Ellington. Indianapolis/Cambridge: Hackett Publishing Company.

Koopman, Colin. 2010. "Historical Critique or Transcendental Critique in Foucault: Two Kantian Lineages" *Foucault Studies*[Online] 8.

Zamora, Daniel. 2014. "Can We Criticize Foucault?" In *Jacobin*, December 10. <https://www.jacobinmag.com/2014/12/foucault-interview>.



cognethic.org