Philosophy in Practice

Volume 9 – Spring 2015

California State University, Los Angeles
Department of Philosophy

Philosophy in Practice

VOLUME 9 – SPRING 2015

Cover Art: Caleb Hahne, “Howling Winds.” 2014. Graphite on Bristol paper.

Contents

Acknowledgments

Philosophy in Practice is produced by students at California State University, Los Angeles. The editorial staff wishes to thank all the authors for submitting and refereeing articles, the faculty of the philosophy department for agreeing to supervise individual contributions, and the University for its generous financial support.

California State University, Los Angeles
Philosophy Faculty

Talia Bettcher (2000–), Chair, Ph.D. University of California, Los Angeles. History of Modern Philosophy, Philosophy of Self, Gender and Sexuality

Mohammed Abed (2008–), Ph.D. University of Wisconsin, Madison. Ethics, Social and Political Philosophy, Philosophies of Violence, Genocide and Terrorism

Bruce Atta (1996–), M.A. California State University, Los Angeles. Metaphysics, Epistemology, Ethics, Social and Political Philosophy

Mark Balaguer (1992–), Ph.D. City University of New York. Philosophy of Mathematics, Metaphysics, Meta-ethics, Philosophy of Language, Logic

Anna Carastathis (2009–), Ph.D. McGill University, Canada. Social and Political Philosophy, Feminist Philosophy, Philosophy of Race, Critical Race Theory, Postcolonial Theory, Phenomenology

Jay Conway (2005–), Ph.D. University of California, Riverside. 19th-20th Century Philosophy, Modern Philosophy, Aesthetics, Social and Political Philosophy

Richard Dean (2009–), Ph.D. University of North Carolina, Chapel Hill. Ethics, Kant's Moral Philosophy, Applied Ethics

Foad Dizadji-Bahmani (2013–), Ph.D. London School of Economics, United Kingdom. Philosophy of Science, Philosophy of Physics, Philosophy of Probability

Ronald Houts (1983–), Ph.D. University of California, Los Angeles. Metaphysics, Epistemology, Logic

Keith Kaiser (2001–), Ph.D. University of California, Los Angeles. Metaphysics, Philosophy of Mind

Steven R. Levy (2005–), Ph.D. University of California, Los Angeles. Epistemology, Metaphysics, Logic, Applied Ethics

Henry R. Mendell (1983–), Ph.D. Stanford University. Ancient Philosophy, History of Ancient Mathematics and Science, Philosophy of Science, Metaphysics

David Pitt (2003–), Ph.D. City University of New York. Philosophy of Mind, Philosophy of Language, Metaphysics

Joseph Prabhu (1978–), Ph.D. Boston University. Philosophy of Religion, 19th and 20th Century German Philosophy, Moral and Social Philosophy, Indian and Comparative Philosophy

Sheila Price (1964–), M.A. University of California, Los Angeles. Recent Philosophy, Comparative Religions, Medical Ethics, Environmental Ethics

Michael K. Shim (2007–), Ph.D. State University of New York, Stony Brook. 20th Century Continental Philosophy, Phenomenology, Husserl, Modern Philosophy, Philosophy of Mind, Philosophy of Language

Kayley Vernallis (1993–), Ph.D. University of California, Berkeley. Moral Psychology, 19th and 20th Century Continental Philosophy, Feminist Philosophy, Ethics, Aesthetics, Gender and Sexuality

EMERITUS PROFESSORS

Thomas Annese (1961–1992), Epistemology, Modern Philosophy

Sharon Bishop (1967–2004), Ethics, Political Philosophy, Philosophical Psychology, Feminist Ethics

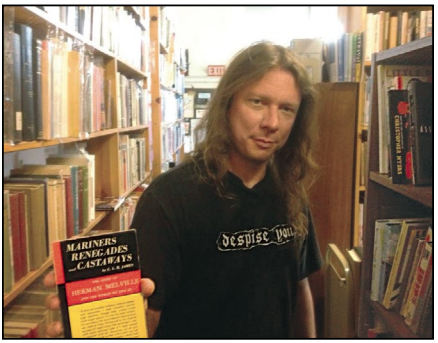
Donald Burrill (1962–1992), Ethics, Philosophy of Law, American Philosophy

Ann Garry (1969–2011), Feminist Philosophy, Philosophical Methodology, Epistemology, Applied Ethics, Wittgenstein, Philosophy of Law

Ricardo J. Gómez (1983–2011), Philosophy of Science and Technology, Philosophy of Mathematics, Kant, Latin American Philosophy

George Vick (1967–1997), Metaphysics, Phenomenology, Existentialism, Philosophy of Religion, Medieval Philosophy

Professor Spotlight: Jay Conway



CSULA and its Philosophy Department have been privileged to have Professor Jay Conway since the fall of 2005. You can't walk the campus with Jay (as he's known to his students) without being approached by multiple students from a variety of disciplines who just want to say "Hi" and share how their lives are going with a professor who clearly made an impact. For many students, his class is the only philosophy class they'll ever take. It's a responsibility that Jay doesn't take lightly. His classes double as writing classes: it's important to him that students take their writing seriously. The goal isn't simply to produce a piece of writing that will earn them the grade they want, but also to produce something they can take pride in creating. Frequently, the students who struggled the most in his classes are also the ones who make the effort to reach out and let him know how much they enjoyed being challenged. This is a testament both to how a university should be and to Jay's dedication to making that a reality.

Jay started his academic career as a Physics major at the University of South Florida. It wasn't until late in his time there that he took a Philosophy class in which they would read Aristotle. He knew that he had to approach writing about Aristotle with a level of detail and precision that he had never previously attempted; the experience was to be a formative one. Needless to say, he never went back to physics. However, in his lectures and writings, you can see the physicist's desire to make dynamic systems comprehensible in language that is both conceptually elegant and informationally rich. Jay went on to get his Master's in Philosophy at USF, writing a thesis on psychoanalysis because,

"it was something I wanted to understand." He completed his Ph.D. at UC Riverside with a dissertation on the French philosopher, Gilles Deleuze, a figure that remains paramount in Jay's conception of philosophy.

While Jay frequently gets lumped into the "Continental" camp of philosophy, he'd likely explain how the Continental/Analytic distinction doesn't really hold up under scrutiny. He then would go on to detail how that distinction functions in different contexts and the possible motivations for making the distinction. His own conception of philosophy doesn't map onto those coordinates. He is interested in the history of philosophy; not as a historian, but as someone who values philosophy and the intellectual impact of engaging with a variety of difficult texts. For Jay, philosophical texts don't represent a line of progress or a record of failed attempts to solve "philosophical problems;" rather, they are living, dynamic entities still capable of doing work. He's just as likely to discuss Plato, Hume, or Spinoza as he is Beauvoir, J.L. Austin, or Deleuze. There is a creative aspect of his conception of philosophy that is clearly visible at the forefront of Jay's research and writing.

Jay brings a level of intensity to the classroom that is unparalleled. You can see him in the library preparing for lecture for hours, and his lecture notes are often longer than the material assigned. He wants to not only analyze a philosophical text with a level of fidelity and rigor that's intimidating, but also to present that analysis with azure clarity and detail. He demands and inspires his students to approach their work in philosophy the same way. Jay never "dumbs things down." He addresses his students as both writers and intellectual equals. He lets you know that: "What we're doing is hard. It's hard for me. And anyone who positions themselves as a person who understands but doesn't struggle with this material is a charlatan."

One student wrote of Jay, "I've taken a course with him every quarter for the past year and a half, and I have continued to keep in touch with him on a regular basis. He is probably one of the only professors that has made me excited about and confident

in my own philosophical career. He exudes a sense of passion for what he teaches, regardless of the specific area of philosophy… He teaches in a way that invites students in, rather than scares them away… He helps students, those that make the effort of course, to realize their own potential. When I first started studying Philosophy I was unsure of my future in it, but after meeting Conway and taking his courses, I am positive that this is what I love doing." Stories like this are common from his students. He wants you to succeed, not just as a student, but also as an intellectual person facing debt, unemployment, and the hardships that any life can bring. It's important to him that his class is a space where you're exposed to difficult ideas and pushed to understand and articulate them. He wants you to learn how to build a relationship with a text that will last after the final exam and your college career; philosophy has a life inside and outside of the university, and he's inviting you to enjoy that life.

There's not really a difference between Jay inside the classroom and outside of it. His office hours used to be at the coffee shop on campus until they tore it down, and he now holds them in the philosophy library. He remembers how intimidating it can be to visit professors in their offices and wants to avoid stifling student engagement. By making himself visible on campus, he effectively engenders more casual opportunities for students to approach him. Every encounter with Jay is an opportunity to do philosophy and he's always seeking to make these encounters be positive and fruitful, for both him and his students. His office hours often turn into conversations between students about class material, making them more of a group help session. It's a space where students feel comfortable conversing with their professor and with each other.

Jay Conway is a treasure to the Philosophy Department and the campus at large. Students and faculty alike have nothing but kind words for him, kind words that he'd be uncomfortable receiving. For Jay, it's his job and his pleasure to do philosophy. He's happy to help.

— M.H. et al

Authenticity, Recognition and Respect: The Power of Dialogue and The Possibility of Failure

Jaewon Choe

In "The Politics of Recognition," Charles Taylor offers an account of Western culture's move from historically class-driven ascriptions of social status to the modern view that all humans deserve equal dignity and respect. This egalitarian view conceives of the status-neutral, rational individual and promotes "difference-blind" equality, rejecting the view that social status and moral worth are dependent upon socially inherited differences. Taylor goes on to describe a modified strain of "equal dignity:" The Ideal of Authenticity and the demand for recognition and respect. Meeting the Ideal requires us to nurture our intrinsic sense of values and asks that we present ourselves in a way that reflects this self-determination. In turn, the demand for recognition asks that others recognize our individual identities in a way that respects their uniqueness. Doing otherwise is to deny the equal worth of our unique identities, effectively objectifying them. Taylor writes:

> The thesis is that our identity is partly shaped by recognition or its absence, often by the misrecognition of others, and so a person or group… can suffer real damage, real distortion, if the people or society around them mirror back… a confining or demeaning or contemptible picture of themselves. (Taylor 1994, p. 25)

In *The Ethics of Authenticity*, Taylor suggests that the Ideal also represents a moral ideal in modernity, treated as a *calling* that ought to be pursued by all (Taylor 1992, p. 17). Within contemporary culture, a tension arises between authenticity and the demands for recognition and respect. On one hand, the Ideal promotes self-

determined identity constructions and resists external influence in their formation. On the other hand, we maintain these authentic identity pictures for higher stakes than mere self-knowledge; we seek their affirmation out in the world. Clinging too hard to the former results in a "soft relativism" where the significance of self-determining identity and declaring, "I am X" has no more value than merely declaring that one has a preference for X. Pushing too hard on the latter threatens the Ideal and suggests that "I am X" is forced on us by the world. Taylor suggests that for an identity claim to be respected, the grounds for making that claim should be demonstrated against "horizons of significance" (Taylor pp. 36-39).

There is also a different tension produced in the modern politics of difference, between the concept of "equal dignity" and our modern strain of differentiated equality. This tension is exemplified in the political debate over affirmative action policies. One side argues that we should judge all persons equally on their merits and disregard social categories like race in order to be fair. The other side replies that disregarding these differences means perpetuating the racist attitudes that have prevented us from living in a fairer society. Each side has reasons for making their respective claims. The former grounds us firmly in universally held human capacities as a means for favoring impartial treatment for all. The latter suggests that denying difference entails a lack of full consideration at least, and potentially oppressive treatment at worst, charging that "neutral" consideration is the tainted ground from which dominant culture has historically imposed oppressive identity pictures onto subordinated groups. As such, many anti-oppression groups work to transform the dominant standards of judgment that produce these bad identity accounts. Just as with authenticity and recognition, extreme forms of one deny the claims of the other.

For Taylor, dialogue plays an important role in easing both tensions. I'd like to explore the power of Taylor's dialogue and apply it to the identity construction of individuals and political interest groups. The aim here is to examine how each construc-

tion reflects what Taylor calls, "the dialogical character" of human experience, and can satisfy the demand for recognition *and* respect. The meeting of the demands comes with a caveat: we must accept the possibility of failure if we are to play this game. At the individual level, identity construction happens in discourse with the world, in concert with horizons of significance. At the group level, politically charged "rebel" collectives aim to transform publicly held, dominant standards of judgment that produce oppressive identity pictures of the collective. In each case, the possibility of failure with regard to recognition and respect is a sign that the activity is taking place on the proper terrain for meaningfully conferring worth: We are engaged in dialogue over identity claims against pre-existing horizons of significance. Without the possibility of failure, we trivialize the significance of self-determining identity at the individual level, and at the collective level, we undercut the fundamental constitution of the collective itself.

"Dialogical" and "Language"

Let's begin by clarifying some terminology. Taylor suggests that full human agency is not determined *monologically* for we do not develop the language of expressing our identities in solitude. Instead, the capacity for self-definition is made possible through the acquisition of this language, which occurs through our dialogue with the world. "Language," as it is used here, is not limited to written and spoken word and instead represents the range of human expressions that shape self-definition, "including the languages of art, of gesture, of love, and the like" (Taylor 1994, p. 33). As such, let's use "dialogue" broadly as well, to capture the dialogical interactions between persons using this broader description of the Language of human expression. The claim here is that, as a range of expressions, the successful acquisition of this language is confirmed by an intelligible demonstration of its use with others. The underlying assumption is that the language of internal dialogue and self-reflection, if it is to be a language repre-

senting human expression over identity, is either demonstrable to others or not a coherent language at all. We should see that our having an intelligible language for use in self-reflection couldn't have been acquired alone. Of course, precluding the role of solitary reflection altogether would be to deny that my living authentically means being present in the world in a manner that *I* deem fit. Taylor suggests that with important issues such as our identity, we construct them through dialogue (Taylor 1994, p. 33).

I'd like to say that solitary reflection has a prominent role, but is merely a part of the larger activity of identity construction. It's reasonable to say that in a culture demanding recognition and respect of unique identities, solitary reflection does real work but only goes so far; we do this in order to present our "true" selves to the world, as a means for reflecting back on the information we've taken in, and we do this for higher stakes than merely holding a landlocked and hidden conception of ourselves. So the dialogical character of this Language acquisition allows us to form ideas concerning identities, and by engaging in further dialogue, we shape our identities in light of those conceptions. In turn, we test these identity pictures in the world, demonstrating that we've accepted, rejected or modified what we've taken in as a means for meeting the Ideal and representing ourselves authentically.

"Horizons of Significance"

So, on what grounds do we analyze what we take in with respect to the Language? On what grounds are we judged by the world in the first place? These determinations may happen within the framework of human expressive language, but how are the values of the language accepted as meaning what we take them to be? In *Ethics*, Taylor talks about a "horizon of significance" (Taylor 1992, p. 39), as the backdrop of intelligibility from which we determine what we value. Just as the successful acquisition of language is outwardly demonstrable, our determinations of value are *respected* only if those determinations are intelligible to others and they concur. In other words, we offer an account of why we

value what we do and present it for public scrutiny. "Intelligibility" seems to assert a ground from which one can judge another's determinations of value and reasonably agree or disagree. The "horizon," then, is a backdrop representing our ideas, feelings, and other sensible things in the world. From this backdrop, we come to value things in various ways. The activity of value-determination may reveal our current standards of judgment to us; shown not only in the objects we value, but demonstrated through the account of how we came to value them. It's possible, then, to see how one's standards of judgment are marked for reevaluation: They may be demonstrated to be lacking in light of the response we receive, and our reflection over the generated response pushes us to reconsider them.

Pictures and Dialogue

Having established the relevant terms, I'd like to color out Taylor's conception of identity construction. Let's think of self-definition as the construction of a picture of ourselves made possible through the employment of the Language. This picture does more for us than merely reflect observations regarding our physical traits and how they're employed. We construct them not just for the sake of self-knowledge over who we are, but we make use of them out in the world. This picture may contain the personality traits that are ascribed to us, our acknowledged preferences, and socio-categorical identifications including race, gender, sexual orientation, ethnicity and so on. We activate various aspects of this picture in some situations, and in others, aspects are activated independent of our will. It's reasonable to believe that, as we move further inward into our social circle, a bigger cut of this picture is recognized. The ideal of authenticity demands that this picture be constructed in a way that aligns with our values and sense of self. In turn, the demand for recognition asks that others recognize the uniqueness of our identity pictures in a way that respects our particular constructions. There appears to be a hierarchal structure to this picture, since it seems inaccurate to say that my preference for

certain things is of equal use to me as my gender or my ethnicity. We should then add that our socio-categorical ascriptions often carry more weight than other, less substantive, aspects of our identities. So the value of my identifying as a heterosexual, Asian male carries more significance to me than, say, my preference for short or tall romantic partners, and even more so than my preference for whiskey over wine. Pushing further here may be mistaken, since it's possible that between individuals with shared identity ascriptions, one may hold gender as primary to her identity while the other does not.

This identity picture, then, is constructed in dialogue with others through employment of the Language. For Taylor, the primary sources for this construction are our "significant others" (Ibid, p. 34). It is through our interactions with loved ones that we first acquire the Language and employ it in shaping our identities. It is in this intimate arena where dialogue produces our first identity ascriptions. These ascriptions may be as innocuous as our being judged as "funny" or "serious," and this is typically the ground where we first learn about more prominent identifications concerning gender and their roles, race, sexual orientation and so on. Let's note here that the construction made in this intimate arena may not always be accurate or even good for us. Taylor says, "It would take a great deal of effort, and probably many wrenching break-ups, to *prevent* our identity being formed by the people we love" (Ibid, p. 34). This is to say that we may form harmful pictures of ourselves, heavily influenced by our significant others. We may come to recognize the harmful nature of these pictures through our judgment of their inaccuracy, but it typically takes much work to undo their significance in relation to how we view ourselves. We should say, then, that whether or not another's judgments are harmful to us, their prominence in forming our identity picture is a sign that they have come from a significant source that we take seriously. In many cases, this seems to explain the harmful effects of *misrecognition*. In this way, we are exposed to, and develop, a horizon of significance through these primary interactions.

Identity as Hypothesis

I'd like to further expand on Taylor's position here. It seems that a fundamental aspect of remixing identity constructions is that we test them in the world, outside of our intimate circle. In a sense, we carry a hypothesis of ourselves, shaped in dialogue with significant others, and present this picture against a broader horizon. Our innocuous identification of being "funny" may come into question when it's shown to us that we are not. Our conceptions of gender roles may be affirmed, denied, or modified in light of what we experience. Our views on race, whether they were accurate or not, get worked on again and again. This, in turn, changes the meaning of these identifications for us, motivating a modified judgment over our relation to them. So we may believe that "being a man" entails 'X(m)' and our relation to it is 'X(m)+'. Here, (m) represents the identity banner of "Man," X represents our currently held descriptions of "Man" and "+" represents our judgments over the banner's descriptive content. When confronted with the information that being a man entails Y(m), we may judge that our relation to this ascription is now Y(m)–. At some point, Y(m) may change to Z(m) and our judgments in relation to Z(m) may change as well. This is overly simplistic. However the idea is that, in many cases, these descriptions, these *metadata*, are where the significance of identity ascriptions lie. Further, their significance is relevant to their use. In other words, it's not just that we seek or reject the ascription of a particular identity banner. Rather, we often reaffirm or challenge the metadata *underlying* the ascription, since that is what determines how we view ourselves, and motivates how we are perceived and treated by others. We may resist these descriptions, but with enough reinforcement, we reflect on what we take in and remix our identity pictures in response. Sometimes our response is a direct rejection of what we take in and we shape our identities in negation to a dominant view. In other cases, it serves to confirm perspectives we've already come to hold. In either case, with enough repetition, certain aspects of our identity take a firmer shape and become harder to deconstruct. In this way,

we present ourselves to the world more confidently over time. We test our identity pictures, take in a response, and reply with a modified iteration that demonstrates our acceptance or rejection of the metadata we are confronted with. At the same time, our new iteration works on the metadata themselves as either a confirmation or challenge of its accuracy.

So we can see that the struggle over identity ascriptions is often a fight over the content of the underlying metadata. This explains why even the more rigid aspects of our identities never quite crystallize, since the meanings of the identifications are open to reevaluation. It's merely less likely, for example, that the meaning of "Man" is easily taken to task, since this identity banner has been determined over a long period of time, and possibly by a patriarchal culture. All this suggests is that over time, experiences and dialogical repetition stabilize identity conceptions. However, as long as the dialogue continues, even the most firmly held aspects of identity are malleable, though it may take more work to question their legitimacy and fracture their alleged authority.

"I AM A 5TH GRADER"

So far, we've colored out identity construction in a manner that appears to rely heavily on external influence and public standards of judgment. At this point, let's consider the following question: How do we construct identity *authentically* if we are so dependent upon external determinations of value? If we are to be the true versions of ourselves in meeting the contemporary Ideal, why should we check our pictures against anything at all? It should suffice to say that "I am X" was determined internally after having acquired the Language. In response, let's consider what Taylor has to say about conferring worth on an identity claim that rejects consideration of external sources. Doing so should reveal how self-determination alone cannot justify the demand for recognition and respect. In *Ethics* he says, "One of the things we can't do, if we are to define ourselves significantly, is suppress or deny the horizons against which things take on significance for us" (Ibid, p.

37). He offers a case where one claims the equal worth of homosexual orientation on the grounds that "those who are inclined to homosexual relations… shouldn't feel themselves embarked on a lesser, less worthy path" (Ibid, p. 37). Now the issue here isn't the determination of sexual orientation, but rather the justification for conferring equal worth. Taylor suggests that in the contemporary culture of authenticity, we tend to grant value to the determination itself, as a means for acknowledging the power of self-determination and the Ideal. It seems reasonable to say, as a moral claim based on equal respect for all persons, that if a person identifies as homosexual that identification should be taken seriously. However, it's still unclear on what account we should judge that such a person is deserving of equal respect. Moreover, how are we to convince those who are morally opposed to the equal treatment of homosexual persons?

This is where extreme self-determination falls flat. In using it as the justification for making the demand for recognition, we trivialize the significance of such an identity claim. In this case, identifying as homosexual has no more significance than saying that we have a preference for "taller or shorter sexual partners, blonds or brunettes. No one would dream of making discriminating judgments about these preferences, but that's because they are all without importance" (Ibid, p. 38). It seems, then, that this extreme form of the Ideal can't have it both ways. If we pursue authenticity as a moral ideal, and we want our identities to be *respected* by others and judged as deserving of equal respect, we must engage with others' judgments. Further, we cannot ignore the fact that pre-existing horizons of significance are already in place. In other words, we require dialogue with others against a preformed backdrop of values. We want to say that the claim "I identify as a 5th grader and am entitled to recess" shouldn't be taken as seriously as "I identify as homosexual and am entitled to marry my same-sex partner," and the tools available for making that discrimination can't be centered around self-determination alone. Finally, we want to say that the first claim is wrong or right for *different* reasons than the rightness or wrongness of the

latter. Each requires its own account from its own context, and we should judge the worth of each account in relation to pre-existing horizons of significance. What we don't want is the conferral of equal recognition and respect hinging on something as perilous as, well, do they *really* identify as such?

Taylor suggests that the equal-value determination of homosexual identity has to be done "more empirically… taking into account the actual nature of homo- and heterosexual experience and life. It can't just be assumed a priori" (Taylor 1992, p. 38). What's interesting here is that this criterion for granting value applies both to those who need convincing and allies who support equal treatment of homosexual persons. In the former case, it's clear that convincing those opposed will require more justificatory ground than self-determined worth. In the latter case, merely granting value as a show of solidarity trivializes the justification for equal treatment. If heterosexual persons can offer an account of rightful treatment on rationally justified intelligible grounds, the account of "homosexual equal worth" should be validated on those grounds as well. Denying that possibility, and confining homosexual identity onto the shaky ground of self-justified rightness generates damaging effects. First, it works to patronize the subordinated group by narrowing the significance of the identity claim. Second, it limits the possibility of convincing those who hold oppositional stances, effectively granting them the firmer deliberative ground.

This is the way in which extreme versions of the Ideal fall flat, trivializing the significance of self-definition and the aspects of our identities that we hold most dear. In a sense, just as our true selves get hidden away by total reliance on others' judgments, the tools for meaningfully conferring worth get left off the table when we reject dialogue altogether. Thus, dialogue is a necessary feature in the demand for recognition and respect. It checks extreme forms of the Ideal in a way that allows us to maintain the value of our authentic identity pictures. For Taylor, our identities are the backdrop from which the things we value make sense. It is because I am *such and such* a person that I hold 'X' dear. If

something I value is made intelligible in relation to a significant other, then that person is part of my identity (Ibid, p. 34). In other words, it's problematic to try and negate the external influences that shape our judgments of value. In a sense, we insist on the value of what we hold dear while attempting to ignore the context that demonstrates their value to us. It would be difficult for one to offer this as an intelligible account to herself, much less offer it outwardly as a means for demanding social recognition and respect. It seems, then, that engagement with horizons is necessary, for better or worse.

The Possibility of Failure

In turn, the dialogical activity of engaging with horizons, self-reflection and remixed response entails the possibility of failure. Extreme self-determination removes the possibility of failure in judgment, since the sources of judgment, in this case, are limited to "I." The removal of this possibility entails what Taylor calls the "narrowing" and "flattening" of the significance of self-determination (Ibid, p. 40). In this way, extreme forms of self-determination, hidden away from public scrutiny, collapse on themselves and cannot meaningfully produce recognition and respect. Denying this is to judge that my identifying as "horse" be taken as seriously as my identifying as "Man." As such, we should not fear the possibility of failure here because it shows us that we're on the right terrain for making our demands for recognition and respect. Perhaps it's a failure of my own judgment and the baselessness of my identifying as "horse." It's strange to acknowledge the possibility that we're wrong about who we are, but it's reasonable to say that at times, our understanding of the metadata that underlie identity banners is mistaken. More significantly, perhaps it's a failure to be recognized properly that stems from faulty *public* standards of judgment, revealing the tainted horizon held up by dominant culture. In this case, we may judge that the metadata itself is inaccurate and our identification as "Man" may be a means of challenging the status quo and highlighting the harmful

effects of inaccurately formed views. As we've seen, making the challenge by ignoring pre-existing standards of judgment is not productive. Rather, what's required is a confrontation with the standards of judgment on the very ground that generated those standards. Put simply, if respect entails more than self-determined respect, we should work to remake the standards of judgment and infuse the pre-existing horizons with more accuracy. Of course, this is no simple task. One way of transforming these standards is to align with like-minded individuals in politically charged collectives.

Rebel Collectives and Resistance to Dialogue

Up to this point, there has been a pushback against self-determination in a way that seems unfair. Even if we grant that pre-existing horizons of significance work on our identities, it hasn't been established that these horizons are accurate, or even good for us. In other words, how should we take the judgments produced in concert with these horizons at face value? In "Politics", Taylor suggests a "Fusion of Horizons" between divergent cultures as a means for recognizing and respecting foreign ways of life. In a multicultural society such as ours, where diversity tangles with the homogeneity of the "founding fathers" conception of equality, there is bound to be opposition to the way identity pictures of historically subordinated groups are presented. In these situations, the dominant standards of judgment that produce such pictures are up for criticism. We see it clearly in feminist and other anti-oppression movements, where part of the struggle for recognition is centered on remaking the descriptions underlying various banners such as "woman," "man," "black," "white" and "privilege." These anti-oppression "rebel collectives" are typically made up of individuals who identify under a common banner and agree that the banner's metadata lacks accuracy, producing oppressive conditions for those falling under that banner. Those who cannot claim the banner but identify with the cause are often deemed "allies."

In many ways, these movements are divergent cultures, created for the sake of challenging and transforming our current standards of judgment. They aim to infuse the broader horizon with more accuracy, in the hopes of producing fair treatment. Sometimes they aim to do away with the pre-existing horizon altogether.

I'd like to examine the formation and function of rebel collectives in order to reinforce the power of dialogue in a collective's meeting a unified, political aim. The formation of these groups traces individual identity construction rather closely. In the way that individual identity pictures are constructed through dialogue with significant others, we can see that the identity of rebel collectives are primarily formed as an intra-collective dialogical activity. While each member's motivation for gathering may be common opposition to something like "institutionalized sexism," and may coalesce in negation to the inaccurate metadata underlying the banner of "woman," how it conceives of "sexism," how it is "institutionalized," and how it should respond in the face of sexist-oppression is determined intra-collectively. In Feminism Is for Everybody, bell-hooks describes the importance of consciousness-raising (CR) groups in the formation of the feminist movement:

> Feminists are made, not born… one becomes a believer in feminist politics through choice and action… before women could change patriarchy we had to change ourselves; we had to raise our consciousness…The (CR) group was a site for conversion. To build a mass-based feminist movement women needed to organize… dialogue was a central agenda at the consciousness-raising sessions…. Argumentative discussion was common in CR groups as it was the way we sought to clarify our collective understanding of the nature of male domination. Only through discussion and disagreement could we begin to find a realistic standpoint on gender exploitation and oppression. (Hooks 2000, pp. 7-8)

This notion of group formation through intra-collective dialogue

shouldn't be a surprise. Any collective of common interests, if they are to have an intelligible identity, requires a constitution of some sort. It's reasonable to see how constitutions of this kind serve as the collective's hypothetical identity; formed through a free exchange of divergent ideas over common concerns. However, let's recall the earlier thought that the formation of an identity doesn't stop at having presented a first account. Rather, it is a continuous process that requires dialogue, self-reflection and remixed response. It seems, then, that while dialogue is important to the formation of a collective's identity, continuous dialogue is just as vital in further shaping and maintaining it. Bell-hooks examines its importance in critiquing the institutionalization of the feminist movement during the '70s and the replacement of CR groups with university lecture halls:

> Once the women's studies classroom replaced the consciousness-raising group as the primary site for the transmission of feminist thinking and strategies… the movement lost its mass-based potential…. The dismantling… all but erased the notion that one had to learn about feminism and make an informed choice about embracing feminist politics… the idea that women needed to first confront their internalized sexism as part of becoming a feminist lost currency. Females of all ages acted as though concern for or rage at male domination…was all that was needed to make one a "feminist." Without confronting internalized sexism women who picked up the feminist banner often betrayed the cause… politicized sisterhood… lost meaning as the terrain… was overshadowed by a lifestyle-based feminism which suggested any woman could be a feminist no matter what her political beliefs. (Hooks 2000, pp. 10-11)

Let's be clear that bringing social movements into the academe needn't entail bad consequences. In most cases, there is a stabilizing effect in having a formalized version of a social movement made possible through direct access to a wider audience. Devel-

oping in the protected setting of academic institutions allows a movement's central themes to coalesce rapidly and integrate into the fabric of society more securely. With that said, there are three points to be made here. First, just as with individual identity, we see the malleable nature of collective identity construction. In this case, we see how the group's fundamental identity shifts with the removal of space reserved for a particular kind of dialogue. As bell-hooks notes, CR groups were places where individuals examined their internalized sexist attitudes as a means for understanding the nature and power of patriarchal society. Without this space for self-reflection, the movement changed its fundamental constitution, for better or worse. What's also interesting is that bell-hooks charges this fundamental change with being responsible, in part, for the media's portrayal of feminism as "anti-male" instead of "anti-sexism." So instead of recruiting male allies, the transformed feminist identity, rooted around "victimization" requiring reparations, produced an anti-feminist backlash from potential sympathizers (Ibid, p. 11).

Here, we see the relationship between collective identity and both intra-collective and outward-facing dialogue. In a sense, patriarchy is a "significant other" to feminist collectives. It represents a primary interaction that produces internalized, harmful pictures of "woman" that work to oppress those falling under that banner. Intra-collectively, dialogue works to clarify the collective's understanding of its opposition and aids in producing a unified constitution for challenging the oppressive status quo. At the same time, the malleable nature of this identity, as a product of intra-collective dialogue, determines the kinds of outward-facing dialogue that are possible. If the identity being presented is perceived as "anti-male," then a different kind of conversation ensues in comparison to an "anti-sexism" identity picture.

The Value of Radical Disengagement

A third point here suggests a certain irony in the function of rebel collectives, and reveals the tension between "equal dignity" and

the differentiated equality generated by the politics of difference. If we seek authentic expression of ourselves, and I see myself as opposing the norms and oppressive pictures that dominant culture paints of me, then why should I seek recognition from such an ill-intentioned force (Appiah 1994, pp. 153-154)? At the collective level, opposition groups occasionally present this view as a justification for total disengagement with previously accepted standards of judgment. Sometimes, they question the validity of the underlying horizons altogether. In some ways, this is an attractive position. To engage in dialogue on this tainted ground is to legitimize the very standards that produce the oppressive identifications in the first place. In this sense, engaging with oppressors cannot complete true transformation, and dialogue is not only unnecessary, it runs contrary to the revolutionary aims of the collective. This radical position is tied to a more reasonable view that the modern politics of difference are a natural successor, and not a particular strain of, the politics of equal dignity. In this space, I'd like to reply directly to the radical view.

As we have seen with individual identity, denying pre-existing horizons of significance trivializes the importance of self-determined identifications. Of course, with rebel collectives, a primary aim of gathering is to object to the oppressive nature of these institutionalized identity pictures and to challenge the current standards of judgment. First, one could ask, just how did this rebel collective construct its oppositional stance and, thus, its identity? If identities are the backdrop from which a collective's dearly held values make sense to them, how should the rebel collective offer an account of what they hold dear without acknowledging the power of the institution they oppose? As a construct via opposition, it seems that rebel collectives cannot disengage in dialogue with patriarchal structure as a means of rebellion, if they wish to maintain the sense in which their values constitute their identity pictures. In other words, without opposition, what does it mean to rebel, and how does a member offer an account of what they value without acknowledging the legitimate power of the patriarchal "significant other"? In seems, then, that the worry over legiti-

mizing patriarchal society by engaging in dialogue with it has no force here. This is because we've already granted its power in our forming the collective in the first place.

Perhaps this response alone is trivially significant and unsatisfactory. Granting that a rebel collective is formed in negation to a dominant view does not necessarily entail that they must engage with it. We could say that disengagement with dialogue changes the intra-collective identity of rebel collectives for the better by liberating its members from the yoke of oppression. Effectively, the collective claims a changed form and has self-determined the worth of the metadata underlying its banner. Having intra-collectively affirmed the value of this new definition, they can now exist peacefully within the confines of the collective. However, what becomes the primary function of this rebel collective? Clearly, the primary aim can't be to transform the public standards of judgment. Further, recruiting like-minded others and transforming the minds of potential allies becomes problematic. On what grounds can this collective convince those who are currently held captive by the dogmatic embrace of the dominant view? Just as extreme self-determination trivializes the significance of individual identity construction, a collective's extreme resistance to dialogue collapses the fundamental purpose of forming a collective in the first place. By our own hand, we remove the possibility of transforming the oppressive standards of judgment that motivated our gathering. Further, we shuck away the potential to transform the lives of others who are equally deserving of a less oppressive situation, merely on the basis that they do not share in our extreme position at a given time. What should we make of them? Is it merely the case that they are not revolutionary *enough* for our tastes?

The most radical stance is to claim that revolution via disengagement is but one move towards the violent overthrow of all oppressive institutions. Even here, if we accept that many will need convincing of this as the *only* solution, and we aren't aiming toward some sort of totalitarian regime of terror, we'll have to offer an account as a means for building a coalition. Whatever

that account may be, if the given reasons justify the stance that non-engagement and force are the only available tools, then the account is evidence of our having acknowledged the power of that which we oppose. In addition, building a coalition that opposes public standards of judgment will have to happen in dialogue against pre-existing horizons of significance. In contrast, if we aim to convince solely through the power of our self-determined position that violence is the only solution, we may not get very far.

In this way, collectives that ignore dialogue as means of protest meet the same fate as extreme self-determinists. It seems that what's left on the table is that we spin our wheels and preach to the choir, as it were, and nothing more. We may intra-collectively determine new metadata for "Woman" for example, but without dialogue the force of the new definition is limited to the confines of our circle. Without the dialogical interplay between oppressive power and determined resistance, we remain bereft of the tools for questioning the authority of the patriarchal status quo and recruiting potential allies. This is because publicly, it may appear that we seek self-determined identity and nothing more. Perhaps that is all that we seek. In that case, we can claim this space for ourselves, but have made a weak case for demanding recognition and respect from others. As such, if we aim to do more than air our grievances, we must consider the conditions under which a productive transformation of oppressive conditions is made possible. If we want our political position to be both recognized *and* respected, and we aim not to destroy society, but to transform it in a way that aligns with our sense of values, the fight must take place on the very ground that produces the oppressive standards of judgment. This is the way we justify the demand for recognition and convince others of the value of our position.

Once again, we are confronted with the fact that engagement with horizons seems inescapable and the entailed the possibility of failure. We may fail to convince others or fail to give our account properly. We may fail our Ideal or be failed by a system that crushes those who oppose it. We shouldn't fret the possibility of failure here. Rather, we should embrace it. This is because, as

Taylor suggests, "Reason is not powerless" (Taylor 1992, p. 41). This arena may not guarantee success, and perhaps it's unfair to be situated in this position to begin with. Nonetheless, this is the struggle. The upshot is that the possibility of failure in this arena entails the possibility of *justified* success, while extreme self-determination and non-engagement enjoy no such significance. This, then, pushes the issue past this ground and generates interesting questions concerning the tension between authenticity, the demands, and conceptions of equal dignity in the political sphere. Does successful change need to happen in our lifetime and does the Ideal entitle us to make this demand? If so, how do we accelerate the process? If radically destabilizing and violent maneuvers are the only accelerant solution, at what point is the potential cost in lives and societal integrity too great to bear? If transformation needn't happen in our lifetime, does this limit the power of authenticity and the demands it can justifiably generate?

Conclusion

There are questions left open here, but I'd like to consider one that can reasonably be answered in the space remaining: What about the possibility of one who is alone in having her correct convictions? Here, one acquires the Language and uses it to show herself how disturbing the current standards of judgment are, correctly observing the mass-acceptance of oppressive identity pictures as norms. The position, so far explained, seems to push back against the possibility that a rebel collective *of one* is justified. However, this view doesn't preclude the possibility of this; it merely limits the probability that one such person is truly alone. So rather than being a party of one, and holding a hidden conception of the world, such a person should seek out like-minded others and collectivize. She should aim to convince others of her position, demonstrating the rightness of her convictions in an arena that contains both the possibility of failure and the potential promise of justified success. What has been shown here is the power of dialogue in both constructing and sustaining identity. For individuals, the process

of presenting identity pictures, accepting a response, and offering a remixed iteration demonstrates the dialogical character of human experience. In political collectives, we see that collective constitutions are formed intra-collectively and tested outwardly against dominant culture via dialogue. In both cases, disengaging in dialogue results in integral losses regarding the significance of our identity constructions. In this way, identity construction doesn't end with our having acquired the Language. Rather, it is a continuous dialogue that consistently works on us, shaping and remixing our identity pictures for as long as we live (Taylor 1992, p. 33). One way to understand this is to say that the mere acquisition of a language doesn't guarantee its appropriate use. Rather, what we should seek in the language of human expression is a *command* of it, as a means for meeting the Ideal. Developing this command requires dialogue, self-reflection and remixed response against a broadened horizon, and this is a pursuit that necessarily shapes itself over time. The acquisition of command, or at least our striving for it, allows us to parse out and reaffirm what we value against these broadened horizons; it puts us in a position to demonstrate mastery over both sides of the interlocution concerning identity construction and the demands for recognition and respect. It is our method for grasping the truth concerning who we are. So, just as we do with other forms of language, we come to understand meaning and significance in an arena that contains the possibility of failure and misuse. We shouldn't fear this possibility. Rather, we should embrace it, knowing we've taken position on the right terrain for producing recognition and respect in a meaningful way.

Bibliography

Appiah, K. Anthony. (1994) "Identity, Authenticity, Survival," *Multiculturalism: Examining the Politics of Recognition*. (Princeton: Princeton University Press)

Hooks, Bell. (2000) "Consciousness Raising: A Constant Change of Heart," *Feminism Is for Everybody: Passionate Politics* (Cambridge, MA: South End Press)

Taylor, Charles. (1992) *The Ethics of Authenticity* (Cambridge, MA: Harvard University Press)

Taylor, Charles. (1994) "The Politics of Recognition," *Multiculturalism: Examining the Politics of Recognition* (Princeton: Princeton University Press)

KANT'S AND RAWLS'S MORAL CONSTRUCTIVISM: TWO THOUGHT EXPERIMENTS

Nefeli Ralli

INTRODUCTION

Rawls introduced the term "constructivism" in the context of modern moral philosophy in his article "Kantian Constructivism in Moral Theory" (Rawls 1980), where he interpreted Kant's moral theory as moral constructivism. According to Rawls morality was either treated as an epistemological problem (as in rational intuitionism) or as a practical problem (as in Kantian constructivism), where moral truth is not something pre-existing, waiting to be found, and where moral principles must be constructed by grounding them in our rationality. According to constructivism, there are no objective moral facts as the intuitionists held. There are no moral truths that are pre-existent and given to us. Instead, moral objectivity or universality should be understood in terms of a constructive procedure that is based on our capacity as rational beings and therefore accepted by all.

Rawls's constructivism rests on the Kantian concept of grounding morality on rationality and impartiality. There is not a Platonic, natural or divine moral order waiting to be discovered. Instead principles stem from the rational choices of moral agents. Furthermore, Kant's Kingdom of Ends, as the ideal state that is based on the categorical imperative, is a thought experiment similar to Rawls's Original Position, in which every individual is a subject and a legislator-sovereign at the same time. The fact that principles are chosen by a hypothetical agreement of idealized agents establishes moral objectivity that is missing from emotivism.[1] At the same time this process avoids the metaphys-

ical commitments of a moral realist, like Plato. However, there are significant differences between the Rawlsian and the Kantian project. The temporal distance and the progress of philosophical thought from the Enlightenment until the late twentieth century led to the revision of the rigorous principles introduced in the *Critique of Practical Reason*. Changes in the metaphysical background and the general methodological framework enabled a better adaptation of morality to the requirements of modern political society. Philosophers have examined the relation between the Kingdom of Ends and the Original Position. Thomas E. Hill Jr. stated that Kant's constructivism represents a fuller comprehensive theory for specific moral cases than Rawls's constructivism. I will argue against it. My thesis is that Rawls was more concerned than Kant with empirical contingencies that affect a person's rational nature and its expression in the social environment, while Kant imagined that we all are equally rational and can easily access our rational aspects expressing our non-empirical perfect rational selves.

In the first two sections of this paper I will examine both the Kingdom of Ends and the Original Position to see how they relate. In the following section I will examine Thomas E. Hill's article "Kantian Constructivism in Ethics," where he contrasts Rawls's Original Position to Kant's Kingdom of Ends and concludes that one cannot simply appropriate Rawls's theory for all purposes of decision-making. I will attempt to demonstrate the opposite. In the last section, I will analyze my thesis that Rawls recognizes that, even if we are rational, we are not all equal and the same in abilities. That is why we have to reconstruct the social structure so that it accounts for the differences inherent in the world and ensures that those differences are not impossible to overcome. On the contrary, Kant considers us as autonomous and rational beings but fails to address the inherent differences that give rise to impediments in our lives. Rawls improved on Kantian constructivism by explicitly addressing the random elements that everyone faces in life, the social impediments and the particularities of every person.

KANT'S KINGDOM OF ENDS

Rawls's conception of an imaginary contract or procedure rests on the Kantian conception of grounding morality on rationality. Autonomy is an important and necessary concept in both Rawls and Kant. In order to see how autonomy is a central concept in Kant's ethics, it is worth examining the categorical imperative: "Act only in accordance to that maxim through which you can at the same time will that it become a universal law" (Kant 1998, 4:421). One is free to act as one wishes but only with the perspective that one's will could become a universal law based on the ground of rationality. Morality is not bound to one's inclination or predisposition; on the contrary, it is *a priori* determined by the authority of the law and the respect owed to it (Kant 1998, 4:426).[2] Practical reason, which is the ability of all people, in general, to answer the question what they ought to do, is not based on any criterion external to itself. It does not employ any content from experience. The moral law holds universal validity, because it stems from the practical reason, inherent in every rational being, not from external psychological and empirical factors. On the contrary, if the imperative is based on an external factor, like personal desires or conformity to principles defined by someone else, then the imperative is hypothetical. Thus, the principle of autonomy is not to choose in any other way than the maxims of one's choice, which at the same time, are universal laws. The moral agent is autonomous because she chooses the principles of her action as the expression of her nature as a free and equal rational being. The categorical imperative, as a self-imposed rule, cannot be proved by the analysis of the concepts of which it is comprised, because it is a synthetic proposition.[3] However, at the same time it must be grasped completely a priori. The moral law is formed through the pursuit of the moral agent to legislate a universal law for the whole society. No contribution is expected from the particular inclinations of human beings, but only from the authority of the law and the respect owed to it. The will of every rational being is a universally legislating will.

Autonomy is the most important element here. One's will is good only when it is an end in itself and not a means to something external. For example, a good-willed salesman will be honest because rationality requires him to be honest, not because he would have a good reputation and gain more clients. According to Kant, there should not be any empirical or natural inclination that motivates one's will except for the good will itself as an end. Thus, human reason is an autonomous source of moral principles, and not heterogeneous and defined by external empirical motivations. A good will leads to the performance of an action as an end in itself, for the sake of fulfilling duty rather than a means to some externally desirable consequences or personal satisfaction. One's duty is therefore categorical and not hypothetical, meaning that one's conformity to the moral law happens for the sake of the moral law itself, not for some other heterogeneous end.

Kant continues by claiming that all rational creatures exist as ends in themselves and not simply as means for the arbitrary use of others. Thus, that formulation of the categorical imperative, which is called humanity as an end, is: "For, all rational beings stand under the law that each of them is to treat himself and all others never merely as means but always at the same time as ends in themselves" (Kant 1998, 4:333). According to Kant, man is subordinate only to the law that he has himself set as an universal law and he is obliged to do so only in accordance with his own will. Every rational being must consider itself as universally legislating through the maxims of its will. The systematic union of several beings legislating as ends in themselves is called the Kingdom of Ends. All rational beings, free from subjective inclinations and predispositions, legislate and obey the universal law they impose on themselves as legislators and equal members at the same time. The maxims produced by this legislating act impose a necessity of action on each member, i.e. duty. Duty rests on the rational will of each legislator who is considered to be an end in herself. The Kingdom of Ends is a hypothetical state, though every rational being must act as if she was a member of this perfect union.

> By a kingdom I understand a systematic union of various rational beings through common laws. Now since laws determine ends in terms of their universal validity, if we abstract from the personal differences of rational beings as well as from all the content of their private ends we shall be able to think of a whole of all ends in systematic connection (a whole both of rational beings as ends in themselves and of the ends of his own that each may set himself), that is, a kingdom of ends, which is possible in accordance with the above principles. (Kant 1998, 4:433)

In the Kingdom of Ends everything has either price or dignity. The concept of dignity is the foundation of duty and it is comprised by the fact that every rational being obeys only that law which she gives to herself. Dignity is contrasted with price. Everything that has a price can be replaced by something else whereas dignity is irreplaceable. Every human inclination, need or delight has a price and a relative worth but dignity has only inner worth. Since morality is the condition under which all rational beings can be ends in themselves, humanity and morality are the sources of dignity. The moral law in the Kingdom of Ends rests on dignity and unconditional worth and it is the source of moral respect. Autonomy is the ground of human dignity.

Rawls's Original Position

On the other hand, John Rawls, one of the most prominent political philosophers of the 20th century, wrestled with the issue of distributive justice. Rawls's main focus was how to construct a just society contingent on his conception of justice as fairness. How can we know which are the principles of a perfectly just society? The answer, for Rawls, consists in discovering the right procedure according to which impartial subjects might construct just principles.

Justice should be understood as that which would emerge as the content of an imaginary agreement made between people deprived of any kind of knowledge that would otherwise make

the agreement unfair. The procedure that leads to the principles of justice is a kind of a bargain made between impartial people—who don't know who they are going to be in the society—and it is called "The Original Position."

Autonomy is also the necessary attribute for the parties in Rawls's Original Position. The parties in the Original Position are considered to be rational, autonomous agents of construction. Further, their deliberations about justice are not bound by any pre-given principles of morality or ideology. In other words, they are not constrained by any principle external to their parties' own rational self-interest.

Rawls lays the groundwork for his famous thought experiment by setting forth an imaginary situation in which rational agents, behind the Veil of Ignorance, are trying to agree on the principles of justice that would create a just and fair society. Imagine, he says, that these people, who are deprived of all the knowledge that might serve to distinguish the one from the other, open a dialogue about building a future just society. Fairness is more easily obtained if they begin from a place of ignorance. The rational agents participating in the discussion know nothing about their personalities, their characteristics and their social position in the future society. They do not know their sex, or their racial and ethnic identities. They do not know if their parents are rich or poor, or whether they belong to an extended, nuclear or a single parent family. They do not know what religious group they belong to or what beliefs, or ideologies they embrace. Above all, they do not know anything about their physical appearance or about their skills or disabilities.

Rawls believes that essential social fairness would be achieved only under these conditions and that the possession of the kinds of knowledge described above would tend to make people partially inclined to various principles. The people in the Original Position are rational. They can reason about and pursue their own interests, but they are not committed to any antecedently given perception of the good or other kind of ideology. However, they have the capacity to pursue and frame a conception of the good,

protect it, and create the best conditions for it to emerge. They are just reasoning about primary goods: liberties, material prosperity, opportunities and equality. They all want these things because the primary social goods will help them achieve their specific goals in the future society. Ignorance of their condition in the future social structure will make them construct unbiased principles about justice that will bind them together in a fair and just social environment. Thus, they will all accept them as reasonable and respect them as objective rules that lead to the existence of a well ordered society.

Rawls's constructivism is considered to be more political than moral. However, in the Original Position the members do not merely negotiate about the principles of fairness and the principles that apply to institutions. Other kind of principles must be also chosen, since a complete theory of right includes principles for individuals as well[4] (Rawls 1971, p. 108). A conception of justice must be accompanied by principles that fall under the concept of right, securing the autonomy and the freedom of persons and setting happiness as the end of a rational plan of life. The principles that apply to individuals are either obligations or natural duties. The principle of fairness is an obligation and holds that an individual is required to do her part by adhering to the rules of an institution, so long as the institution is just. On the other hand, natural duties are positive or negative. The positive duties are the duties of mutual aid and respect and the negatives are the duties not to harm or injure another and not to cause unnecessary suffering. The usage of the word "natural" implies that those duties hold between persons independently from their institutional relationships. They apply to equal moral persons. Therefore, it seems that Rawls's theory can be considered as moral constructivism. The notion of fairness is wider than the distribution of justice or social justice. It includes the notion of rightness and morality.Rawls was concerned about issues such as the foundations of moral knowledge, the role of rationality in the procedure of deciding the principles of justice and he was critically opposed to utilitarianism.[5]

Both the Kingdom of Ends and the Original Position are

hypothetical ideal states and operate as heuristic devices that ground morality on rationality. The rational agents in both thought experiments are idealized, autonomous agents and they abstract from their personal differences. In both theories moral principles are the objects of rational choice and thus exclude any kind of heteronomy on the decision-making. Rawls adds to this concept that the principles chosen must apply to the basic social structure by introducing the Veil of Ignorance as a necessary condition. Rawls adds to the categorical imperative the condition that the agent is ignorant of her future social position and has to choose principles that promote equal opportunity for every possible member of the society. That is the reason why Rawls thinks that the Difference Principle would be an important product of the agreement derived within the Original Position: social and economic inequalities are to be arranged to the greatest benefit of the least advantaged.

Criticism of Rawls's Constructivism

Thomas E. Hill Jr., in his article "Kantian Constructivism in Ethics" (Hill 1989), thoroughly examines the two thought experiments. He states that Rawls's theory as it stands is inappropriate for all purposes of moral decision making. The Original Position does not offer sufficient yardsticks for individuals facing immediate moral choices and cannot be used to address matters beyond the justice of basic institutions. According to Hill, the procedure is also limited to an "overlapping consensus" concerning principles that stems from Western democratic traditions. Hill uses a specific example in his effort to show that Rawls's approach does not offer a full comprehensive theory for moral choices faced by rational agents in concrete situations. Imagine, he says, that I am dealing with a dilemma whether to lie to a friend, Carla, about the past infidelities of her recently deceased husband, Floyd. Hill concludes that, in this specific situation, I would not be able to address my problem directly from the perspective of the Original Position, behind the Veil of Ignorance, since the primary goods are the exclusive focus of agents in the Original Position. I would

have to put aside my sympathy, my particular commitments to my friend and disregard any knowledge of the special things we value as ends. As Hill states, the Veil would blind me to potentially relevant facts about personal relationships (Hill 1989, p. 760).

There are two key objections to the previous argument. Firstly, there is no reason to restrict Rawls's moral guidelines to the principles chosen in the Original Position. In the Theory of Justice, Rawls explicitly explains his view. In the Original Position the concept of goodness is explicated only in a rather thin sense. For Rawls, the concept of right is prior to the concept of good. Something is good only if it fits with the principles of right and it is determined by a person's rational plan of life in an Aristotelian sense. The thin theory of good that is dominant in the Original Position secures the principles about the primary goods. When the principles about justice are secured, the Veil is lifted and we can develop the first principles into a full theory of good. The definition of good, according to the full theory, applied to simpler cases is: A is a good X if and only if A has the properties which it is rational to want in an X, given what X's are used for (Rawls 1971, p. 399). That is an Aristotelian view that entails a functional concept of goodness that is discovered by one's rationality. Good is defined in terms of a goal for which something is made or that towards which it moves, and rationality is defined as taking effective means to achieve one's ends. Additionally, Rawls states that since the Veil is lifted the basic value judgments are those made from the standpoints of persons, given their interests, abilities and circumstances. There is nothing a priori about moral philosophy and moral judgments are dependent upon social circumstances.

The principles of justice derived from the Original Position prepare the way for extending the definition of good to larger questions of morality. Thus a good act is one which we are at liberty to do or not to do, when no obligations or natural duty constrain us, and when the act leads us to the advancement of another's rational plan (Rawls 1971, p. 438). In response to Hill's challenge, would it help Carla to know about Floyd's past infidelities? If Carla is a person that desires to know the truth and expects her friend to be

honest, then her friend should tell her the truth. In the Original Position the first principles single out relevant features of moral situations such that a particular exemplification of these features could provide a reason for making a moral judgment. Thus, while behind the Veil, it might be difficult to use moral guidelines to derive imperatives for conduct in individual circumstances, as Hill states. However, after the introduction of the full theory of good moral guidelines for specific moral cases can be found.

Secondly, we could extend the principles of the Original Position to broader ethical considerations and use this expanded set of principles to supply an answer to Carla's friend's moral dilemma. Besides the first principles of justice, the rational agents behind the Veil, while in the Original Position, determine some obligations and natural duties of mutual aid and mutual respect. Thus, a rational agent could possibly be led to the conclusion that since she wants to promote a rational plan in her life (whatever that might be) she does not want to be deceived or to deceive because this would hinder the development of her full potential. Knowing the truth always helps one to understand one's condition better. Mutual aid and mutual respect as natural duties entails that individuals do not want anyone to be deceived because deception is not compatible with respect.

Hill also criticizes the fact that members in the Original Position are mutually disinterested, meaning that they take no interest in each other's interests and thus have no other-regarding motives moral or otherwise (Hill 1989, p. 764). Hill states that this could lead to amoral egoism, since the members of the Original Position do not adopt any pre-given moral principles and they are concerned exclusively with their own self-interest. Grounding the construction of principles of justice on self-interest without particular knowledge is really convincing when determining the concept of fairness, but it would not help deliberating on a wide range of moral and interpersonal issues.

However, Hill's criticism can be refuted. To think of Rawls's theory as a theory of amoral egoism is misleading. The fact that the deliberators in the Original Position are mutually disinter-

ested does not entail that, in everyday life, they are not interested in one another. The two Principles of Justice—the principles of obligation and natural duty derived from the Original Position—demand the mutual respect and concern for others (Rawls 1971, p. 148). Mutual disinterest as a motivational assumption accords with Kant's notion of autonomy because it allows freedom. Rawls, tried to avoid any ideological restriction on the parties of the Original Position. If they were considered to be altruists their choice would be confined to principles compatible with altruism. The parties in the Original Position only desire certain primary goods because it is rational to desire them in order to fulfill any kind of rational plan. However, even in the Original Position the combination of the Veil of Ignorance and mutual disinterest leads to the same advantages as benevolence. Being deprived of knowledge about their future condition and remaining mutually disinterested, the deliberators would take into account the good of every member of the future society. If a person deliberates about fairness from a neutral standpoint, knowing that she could be anyone in the future society, she would care about the welfare of each future member of that society. The claim that the conception of the Original Position is a form of amoral egoism could be true if we take into account only one aspect of its components, the mutual disinterest. Nevertheless, the assumption that one could construct moral principles on the grounds of benevolence, full information and impartiality would lead to utilitarian conclusions and would make any initial agreement on principles rather difficult.

Furthermore, Hill considers that every theory that deprives the rational agents in the Original Position of historical and cultural information renders any guidance to all morally relevant circumstances difficult. Principles that object to killing and bodily injury can be derived under all conditions, but disagreements arise about how to make such principles specific. According to Hill, the articulation of moral principles may vary to some degree with historical and cultural conditions (Hill 1989, p. 763). Thus, we cannot disregard the cultural context of society when we are deliberating about principles.

However, Rawls does account for this problem. The parties in the Original Position do not know particular facts concerning the economic or cultural circumstances of their own society. But they know general facts about human society and they understand political affairs and principles of economic theory. They are presumed to know all the general facts that could affect their choice. According to Rawls, the parties are deprived of information, but they have enough knowledge to rank the alternatives and they know which situations to avoid (Rawls 1971, p. 143). Thus, if they have general knowledge about facts, the parties would agree on the best possible principles from a wide variety of alternatives. They would agree on principles that would be equally fair when applied to different cultural contexts. Furthermore, one person's rational choice does not depend on how much she knows, but on how well she can reason using whatever information is available to her (Rawls 1971, p. 397).

Comparison of Rawls's and Kant's Constructivism

Comparing the Original Position to the Kingdom of Ends, Hill concludes that Kant's theory deals more coherently with general moral guidelines that can be applied to specific situations. Firstly, the Kantian legislators have values beyond the primary social goods and they are committed to the value properties that are essential to being rational because they consider rational nature as an end in itself. Secondly, according to Hill, the Kantian legislators are not mutually disinterested. They are committed to furthering the ends of others because they consider the dignity of each rational agent to be the highest value. Finally, the Kantian formulation of the categorical imperative abstracts from all differences between rational agents and disregards historical and cultural differences. But, as Hill states, this is not a problem for the Kantian model because it represents a fuller moral comprehensive theory than Rawls's theory. The categorical imperative leads rather directly to general moral guidelines. The Kingdom of Ends promotes a strict

principle of respect for persons, an indefinite principle of beneficence, the consideration of the ends of others, and treating others as ends in themselves (Hill 1989, p.770). However, as I tried to show previously, if we take into account Rawls's full theory of good, we are provided with moral guidelines for general purposes. Concerning the objection from mutual disinterest, I argued that his theory is not a form of amoral egoism. In the Original Position the combination of the Veil of Ignorance and mutual disinterest leads to the same advantages as benevolence.

The main difference between Kant's and Rawls's thought experiments lies elsewhere. In particular, the difference lies in Rawls's deepening of Kant's conception of choosing on rational ground by adding the Veil of Ignorance. He adds the requirement that the principles chosen are to apply to the basic structure of society and the principles derived should be chosen in the light of natural restrictions and the random contingencies that are inherent in our world. When one acts as morally required, one is expressing acceptance of the rules for society that take account of the potential for inequality and unequal development.

Kant states that man realizes his true self when acting out of respect for the moral law, whereas if he acts on sensuous desires or other situational contingencies he is no longer autonomous. However, Sidgwick criticizes this idea, pointing out that the lives of a saint or a scoundrel could equally be the outcome of a free choice on the part of the noumenal self and equally the expression of the phenomenal self.[6] Kant does not explain why the life of a scoundrel does not express his freely chosen selfhood in the same way that a saint actualizes his selfhood in living the life of the saint (Rawls 1971, p. 254). Kant stresses that one can choose a consistent set of principles as a noumenal self, and that acting from those principles is an adequate condition that renders one a free and equal rational being. One might say of this point that Kant believes that we express our freedom only when acting rationally. But, according to Rawls, Kant does not give enough reason to support the idea that acting from the principles of the moral law expresses our nature in an absolute way, while acting from

contrary principles does not (Rawls 1971, p. 255).

In contrast, by introducing the concept of the Original Position, Rawls tries to bridge the gap between the noumenal selves and their expression in the actual context of society. The Original Position is the starting point that shows which principles would be chosen by any free and equal rational persons that must be applicable in practice chosen in the light of natural restrictions. At the same time these rational persons declare their independence from the contingencies of nature and society. As noumenal selves, the parties choose the principles that would manifest their freedom in the actual community and ensure their independence from any natural and social accident. According to Rawls, these principles reflect the moral law that stems from the nature of a free and equal rational being. As Rawls says "Our nature as such beings is displayed when we act from the principles we would choose when this nature is reflected in the conditions determining the choice" (Rawls 1971, p. 256). This nature is not the nature of a transcendental being, but the nature of a rational being whose life is determined by the random social and natural lottery. The notions of autonomy and rationality are not purely transcendental, nor are they devoid of connections with human life and contingent social condition. The members of the Original Position acknowledge that dependence. The persons in the Original Position stand as noumenal selves but cooperating as a collective force. The choice of oneself, as a rational and equal human being, must be seen as a choice acceptable to other selves. But above all, it is the choice of beings that know that are going to be subjects to the conditions of human life. The freedom of pure intelligence, independent of the restraints of the natural lottery does not concern Rawls's theory of justice.

For Kant, social situation is less important. Kant would exclude any kind of rational decision depending on the avoidance of natural and social contingencies as hypothetical, as a choice not depending on free practical reason but to a certain end. Practical reason, for Kant, generates its own unique motive and a rational will should be understood as distinct from a will tied

to social necessities because such a will would be heteronomous. Action arising from respect of the moral law is the only action that expresses any true concern for morality. Thus, the concept of freedom of rational agents is grounded in the detachment of the categorical imperative from social circumstances. Kant presents the moral individual as superior to any actual social order.

For Rawls the idealization of the rational deliberators is done for a specific purpose. This is to exclude the imperfections and also the empirical differences in circumstance between real humans who might bargain. Thus, the idealized rational nature of the parties in the Original Position leads to principles that will be generally and objectively accepted. For Kant, the idealized rational nature of oneself is always accessible to oneself and capable of constructing a universal moral law.[7]

Rawls recognizes that, even if we all are rational, we are not all equal in abilities. Specific conditions in life lead us to different kinds of rational choices. As a result, we have to reconstruct the social structure so that it accounts for the differences inherent in the world and ensures that those differences are not impossible to overcome. Justice is not getting everything one wants but equal opportunity to achieve one's goals. Rawls acknowledged that a just society wants to promote the intellectual and moral development of citizens, for the sake of the stability of that society. This way he gives more attention to empirical contingencies that affect a person's rational nature and its development. On the contrary, Kant considers us as autonomous and rational beings but fails to address the inherent differences that give rise to impediments in our lives by focusing on the importance of the perfect rationalized version of ourselves. One might argue that the categorical imperative implicitly accounts for differences but these differences are too important not to be explicitly addressed.[8] Rawls improved on Kantian constructivism by explicitly addressing the random elements, social impediments and the particularities of circumstance that everyone faces in life. So, it seems that, a system of morality and justice should take into account differences and inequality that inhibit a person from acting the same

way as another person in a more advantageous position.

Conclusion

Moral Constructivism denies the existence of pre-given moral properties, and stresses the necessity of constructing moral principles by grounding them on rationality. I explained aspects of both Kant's and Rawls's constructivism, examined their similarities and discussed the two thought experiments that are grounded on the conception of autonomy. I have tried to show that Rawls's constructivism can be offered as a guideline for more specific moral judgments if we take into account his full theory of good after the Veil is lifted. I also tried to show that Rawls's element of mutual disinterest does not lead to amoral egoism, but functions as a starting point for moral neutrality. Combined with the Veil of Ignorance it does not lead to egoistic motivation but rather to collective self-interest that necessarily ensures equal opportunities for everyone in future society, because it presupposes that the principles derived are chosen by each individual as an equal member.

I have argued that Rawls insists that the members of the Original Position should be ignorant about the particulars facts about their lives because he accounts for the differences inherent in the world and ensures that those differences are not impossible to overcome. Kant considers us autonomous and rational but fails to address explicitly the inherent differences that give rise to impediments in our lives, rather focusing on the perfect rational aspect of our nature. Rawls includes in his constructivism the ties of every rational person to an actual social order.

Notes

1. Emotivists believed that moral sentences do not correspond to true or false propositions but they express moral attitudes.
2. A priori means independent of experience and even of all impressions of the senses.
3. The contrast between synthetic/analytic propositions was introduced by Kant. Analytic proposition is the one where the concept of the predicate is

contained in the concept of the subject. A synthetic proposition does not include the concept of the predicate in the concept of the subject and is liable to substantial information.

4. Rawls does not use the concept "right" as it is used in its traditional sense. According to him the broader notion of rightness as fairness is to be understood as a substitute for the existing terms. Thus, Rawls's notion of right is also a case of constructivism, where the acceptance of a pre-given concept is excluded.
5. Rawls criticized the utilitarian view for adopting the stance of an impartial spectator of society and adopting for society as a whole the principle of rational choice of one man, thus utilitarianism does not takes seriously the distinction between persons (Rawls 1971, p. 27). It allows the sacrifice of a few for the greater good of many.
6. Noumenon is a term associated with Kant referring to things as they are in themselves as opposed to things as they are for us knowable by senses (phenomena). Kant stated that we need to postulate a noumenal reality and a noumenal self as a condition of free will that is opposed to the phenomenal self who is being determined by her actions.
7. Special thanks to Professor Richard Dean for suggesting this point.
8. When one tests a created maxim to see if it could be made universal law, one is asked to imagine if the law is fair no matter what one's station in life is.

Bibliography

Hill, Thomas E. Jr. (1989) "Kantian Constructivism in Ethics", *Ethics*, Vol. 99, No. 4, pp.752-770

Kant, Immanuel. (1998) *Groundwork of the Metaphysics of Morals*. Mary Gregor and Jens Timmermann (Ed). (New York: Cambridge University Press)

Rawls, John. (1971) *Theory of Justice* (Massachusetts: Harvard University Press)

Rawls, John. (1980) "Kantian Constructivism in Moral Theory", *The Journal of Philosophy*, Vol. 77, No. 9, pp. 512-572

Reasoned Disagreement in The Public Sphere

Austin Mitzel

In his article entitled "The Idea of Public Reason Revisited," Rawls claims that the idea of public reason belongs to a conception of a well-ordered constitutional democratic society (Rawls 1997, p. 573). He is not alone in his sentiment. Many theorists of democratic liberalism, concerned with the protection of citizens from undue interference by the state, have articulated conceptions of reason which belong properly to the political domain and successfully distinguish public reasoning from debates concerning comprehensive doctrines and other private moral questions. This theorizing has raised the question of the nature of political disagreement and its relationship to social and cultural differences. Do the religious, ethnic, and cultural differences between citizens abrogate the usefulness and productivity of political deliberation? Do these kinds of differences have to be laid aside when entering into the public forum? Rawls argues that a well-ordered democratic society calls for a kind of reason specifically appropriate for the public domain—a reason to which all citizens, as citizens, can appeal when engaged in public deliberation. I will argue in this paper that his concept of public reason is, in some instances, insufficiently robust and cannot arbitrate in key political debates. In other ways it is too restrictive and unwittingly excludes marginalized groups and individuals from political deliberation. I will then offer a modified notion of public reason which I think successfully meets the challenges to which his conception is susceptible.

My paper will be structured as follows. First, I will briefly characterize Rawls' notion of public reason and its relation to Rawls' other political concepts. Secondly, I will offer a challenge to his notion of public reason. By way of a debate currently

raging in the United States concerning education, I will attempt to show that the sole use of public reason, though sufficient in many instances, would fail to resolve all political disputes. As a result, practical resort would be made in such cases either to non-deliberative democratic procedures or to non-public grounds of agreement. I will say something about why both of these are undesirable alternatives. The first is unacceptable because, in the absence of public reason, citizens would be unable to justify their positions to each other and would therefore have no reason, aside from fear of legal retribution, to regard various political outcomes as legitimate. I will then argue that the second alternative would likely result in the exclusion of certain groups and individuals from political deliberation. Finally, I will offer a modified notion of public reason which meets my argument from education and would support the implementation of a more inclusive model of democratic deliberation.

Liberalism and Public Reason

There are good reasons for abandoning the idea that the state should promote a specific idea of the good, or a comprehensive doctrine. History seems to show this. Consider the religious conflicts of the 16th and 17th century in England, France, and Germany. Untold lives were lost to state intolerance. Many killed in support of the idea that a specific religious or denominational ideology had to be promoted and embraced by the state. One outcome of reflection upon these conflicts was the formulation of a political liberalism and the articulation of a family of notions of individual rights. On many of these systems[1], the notion of the government as a repository for robust comprehensive doctrines is rejected in favor of a picture on which the government provides a political framework in which citizens are free to order their lives as they please in accordance with a set of rights and liberties. A broad private sphere is distinguished from the public sphere, and one's choice of comprehensive doctrine is relegated to the private sphere.

Since in liberal theory the role of government is limited to

guaranteeing the rights of its citizens and providing other basic public goods, its success as a theory depends upon a clear demarcation of the public and private realms. For Rawls, one of the most famous liberal theorists of the 20th century, the legal recognition of public and private spaces is insufficient. A satisfactory theory of liberalism must also articulate distinct forms of public and private reason corresponding to the public and private domains. What is the content of these domains? In the private sphere, individuals are free to embrace all kinds of mutually inconsistent comprehensive doctrines, beliefs, cultural affinities, religious ideas, ethnic identities, etc. They are free to act on these private conceptions so long as their actions do not prevent others from gaining access to the goods which they are guaranteed in the public sphere (including the ability to embrace various other kinds of comprehensive doctrines, beliefs, etc.). To the extent that they recognize the same rights for others which they would reserve for themselves, liberalism recognizes all such conceptions as rational. The content of the public sphere, on the other hand, is given by the primary social goods—those goods which Rawls thinks everybody may reasonably be thought to want, apart from whatever other private interests they may have: rights, liberties, opportunities, income and wealth, and the social bases of self-respect (1975 Rawls, p. 260).

Rawls elucidates the distinction between the public and private spheres with what he calls the Original Position. The Original Position is a theoretical apparatus by which public interests can be distinguished from private ones. The Original Position describes a hypothetical situation inhabited by a group of legislators who are asked to articulate the basic concepts and principles of justice which ought to be embraced in a society on whose behalf they deliberate. Crucially, those who deliberate in the Original Position do so behind the Veil of Ignorance—that is, without any particular knowledge of the country, population, cultural affiliations, religions, or social standing of those on whose behalf of whom they deliberate. It is a thought experiment designed to elicit the principles of justice that would be chosen from within a situation of neutrality. Because those in the Original Position lack

particularized knowledge, their opinions cannot be influenced by the private interests of those that they represent. Rawls thus reasons that the scope of deliberation within the Original Position would be restricted to the primary social goods—the goods which "Rational persons may be presumed to want whatever else they want" (Rawls 1968, p. 158). Since the primary social goods are those which any rational person would want regardless of whatever private ends they may have, the government can justly distribute these primary social goods without privileging one conception of the good over another—hence, without violating the principles of liberalism.[2]

When the legislators deliberate from within the Original Position behind the Veil of Ignorance, they are unable to appeal to private interests or conceptions of the good. Their deliberation is thus restricted to the primary social goods, since these goods alone are desired whatever other ends a rational person might have. This indicates the basic distinction between public and private reason. Public reason has as its content public interests—namely, the just distribution of primary social goods. When engaged in public deliberation, citizens cannot advance exclusively private interests—those interests that lie outside the purview of the primary social goods. The sorts of reasons to which citizens can appeal are restricted to those concerning public interest. That is, when citizens deliberate concerning policy, they must appeal to reasons which others citizens as citizens[3] might accept. The reasons given in public discourse must not be exclusively linked to specific comprehensive or private conceptions of the good. Rather, they must satisfy what Rawls calls the criterion of reciprocity. The criterion of reciprocity is satisfied "Only when we sincerely believe that the reasons we would offer for our political actions… are sufficient, and we also reasonably think that other citizens might also reasonably accept those reasons" (Rawls 1997, p. 578).

One might be inclined to think that deliberation in accordance with public reason necessarily ends in consensus. Since, when using public reason, citizens only appeal to principles and values pertaining to the primary social goods, and give reasons

that other citizens, as citizens, might accept—where might disagreement arise? This view is mistaken. There are two sources of disagreement that remain within the sphere of public reason. First, even if all citizens authentically engage in public reason, they are still subject to certain epistemic limitations. Perfect understanding of all the relevant facets of a piece of legislation is impossible. Often, the complexity of political proposals results in disagreement about whether or not they advance public interest. Disagreements of this kind are essentially interpretive—disagreements in which all participants agree on a set of values or goods, which ought to be advanced, and disagree about whether a given piece of legislation successfully advances them.[4]

But there is a more significant kind of disagreement that can occur in the realm of public reason. This is due to the fact that several forms of public reason are realizable within a liberal democratic regime. All legitimate forms of public reason satisfy the criterion of reciprocity.[5] This considerably narrows the field of possible forms of public reason. Forms of reasoning based upon religious affiliation, for example, do not belong to public reason. This is because one cannot expect another person qua citizen, being interested only in the advancement of certain public goods, to accept such reasons as legitimate. A form of reasoning satisfies the criterion of reciprocity when the reasons comprising that form are thought to be reasonable[6] to others—reasons that others, as reasonable deliberators, might accept.[7] There will, therefore, be some overlap in terms of the interests advanced in each form of public reason. Rawls thinks that any legitimate form of public reason will involve some rights and liberties, and perhaps the advancement of other social goods. Kinds of public reason merely differ with respect to the exact specification of the rights, liberties, and other social goods on which they are based.

Both of these kinds of disagreement, for Rawls, are a natural occurrence within public deliberative discourse. They are not, therefore, inconsistent with the basic distinction, central to liberalism, between public and private reason. Disagreements arising from epistemic limitations concerning policy and varia-

tions in conceptions of liberalism still belong within the public realm. Those disagreements consistent with Rawls' concept of public reason will concern the primary social goods and the kinds of reasons which might be acceptable to anybody qua citizen. In the public sphere, people disagree as citizens, not as members of cultural associations, religious organizations, or ethnic groups. This kind of disagreement is to be distinguished from disagreement concerning the private non-reciprocal beliefs and ends of individuals.

The question which now confronts us concerns the extent to which public reason can be used successfully to justify choices of public policy, and private reason can be kept within the private sphere. Can disagreement within the context of political deliberation be limited to the kinds of public disagreement outlined above? In the next section, I will try to complicate the relationship between democracy and public reason. In particular, I want to argue that there are unavoidable political decisions concerning which public reasoning about the primary social goods is an inadequate support. There are situations in which arguments for alternatives in the political arena cannot be resolved through appeals to public reason alone.[8] I will substantiate this claim in an argument from education.

Non-Public Values in Education

We have seen that, on Rawls's account, a variety of incommensurable conceptions of the good are consistent with basic public values. For the purposes of my argument, I will call a 'thin' conception of the good one which promotes the just distribution of the primary social goods and is neutral with respect to non-public values. Though Rawls does not explicitly list education among the primary social goods, he thinks that education is an ineliminable service that the state must provide:

> In order, therefore, to apply the notion of pure procedural justice to distributive shares it is necessary to set up and to administer impartially a just system of institutions… the

> intuitive idea is familiar. Suppose that law and government act effectively to keep markets competitive, resources fully employed, property and wealth widely distributed by the appropriate forms of taxation… assume also that there is fair equality of opportunity underwritten by education for all. (Rawls 1999, p. 87)

Equality of opportunity is key, for Rawls, in the establishment of a just society. This implies that one of the just actions of the state consists in correcting the morally contingent distribution of social and natural inequalities (Ibid, p. 15). Education is one way in which this can be done. Furthermore, it is one of what Rawls calls all-purpose means—means which can be directed to any end or comprehensive doctrine which a person chooses or believes, so long as that end is consistent others having equal liberty, opportunity, and access to the other primary goods (Ibid, p. 93). Because the primary social goods are all-purpose means, Rawls thinks that a society can distribute these primary social goods without promoting any particular conception of the good. We now need to inquire whether education can be reasonably construed as being neutral with respect to the ends, values, or reasons lying outside the thin conception of the good.

It need not be the case that education is completely devoid of values. In the same way that a regime can never be completely neutral with respect with to goods, an education is necessarily accompanied by the inculcation of values. It need only be the case that the values are the values explicitly entailed by the thin conception of the good. Education in a liberal state will, for example, involve passing democratic values on to future generations, since such a regime has an interest in perpetuating democratic principles. Among Rawls's primary social goods are liberty and opportunity (Ibid, p. 92), so it will be the case that indoctrinating students with the value of liberty and opportunity will feature in this education. If the education provided is successful, students will come to believe that all people should have equal rights and opportunities, and that it is important to respect the liberty of others.

Thus, one can formulate an educational policy that is neutral with respect to conceptions of the good, as I am conceiving of it, while still emphasizing certain public values. The question whether education is totally value-free is uninformative. It is far more interesting to ask whether decisions concerning what ought to be taught publically can always be made exclusively on the basis of public reasoning about the thin conception of the good. My contention is that this is not the case. There are academic subjects and positions within those subjects the teaching of which is commonly accepted to lie within the purview of the responsibility of public education, but which are not entailed by the thin conception of the good. I will give one example. It is commonly accepted that science is one of the core subjects which ought to be taught publicly. Evolution is now widely taught (and almost universally accepted) within the biological sciences. Yet, in the United States, there is significant opposition to evolution. Young Earth Creationism (henceforth YEC) is one thesis which has been widely adopted within certain Christian circles. YEC teaches that the earth is far younger than evolutionists have claimed—somewhere between 6,000 and 10,000 years old, and that it's coming into being was caused by God, who created the earth in accordance with the creation story found in Genesis 1–2.

I think that a case can be made for the claim that YEC is unscientific. It offers no predictions and is untestable. Even where evidence seemingly contradicting YEC has been presented, the creationist can (and has) appealed to the divine will to explain it away. Any idea which has a mechanism that can defuse all possible evidence to the contrary is in principle unfalsifiable. By many standards, then, YEC is not scientific. By contrast, many claim that the teaching of evolution would be more likely to encourage genuine scientific values: it is consistent with and emphasizes the importance of empirical evidence, predictions, and testability. Most importantly, insofar as they are consistent, proponents of evolution agree that, as a scientific theory, evolution might be overturned in favor of a superior theory.

There is decisive agreement within the scientific commu-

nity that evolution is theoretically superior. However, the agreement of a few (or many) as to the relative merits of evolution over YEC does not imply that the values with which it is consistent are entailed by the thin conception of the good. How is it that teaching YEC is not consistent with the promotion of political liberty, and with the just distribution of primary goods? It may be a bad pseudo-scientific idea, but it is not inconsistent with Rawl's thin conception of the good. There may be great value in aligning one's beliefs with empirical evidence and formulating hypotheses that yield predictions, but this is not a necessary corollary of the thin conception of the good or of public reason. It seems that this is enough to challenge the idea that the dominant biological curriculum is neutral with respect to non-public goods. If we choose to teach evolution in the classroom, we must justify that choice by appealing to reasons which lie outside the purview of public reason.

On the other hand, we remain free to teach both evolution and YEC in the classroom. Instead of asserting the theoretical superiority of evolution, one could present both, along with their support and implications, and let the students decide. But it is not obvious that this is desirable course of action. First, this does not seem to accord with the nature of education. It is true that education, in some respects, teaches competing theories and expects students to make decisions based on the evidence presented. But this is not how all or even most education works theoretically or in practice. Particularly in the sciences, education involves presenting information as factual. It is not a process of presenting a matrix of possible choices. Second, presenting evolution and YEC as equally legitimate possibilities might be intellectually spurious.[9] The vast majority of biologists consider YEC to be a pseudoscience. What becomes the state of biology in a country where an idea considered aberrant and unscientific is presented alongside a dominant scientific paradigm? Third, the presentation of YEC alongside evolution raises difficulties concerning other kinds of theories. If a third theory of biological origins was presented, it seems to be the case that it would need to be displayed along-

side both evolution and YEC. If not, a principled reason would have to be supplied as to why it is being excluded. But no matter what criterion we choose, some theories would necessarily slip through. Consistent application would then result in the proliferation of theories being taught, and biology would become an impossible subject to teach.

I think that this shows that there is no obvious course of action concerning the educational policy in question. The content of the curriculum being debated is underdetermined by the thin conception of the good. The choice of teaching biology in accordance with evolution, to teach both YEC and evolution, or to pursue some other alternative requires appealing to values that lie outside the thin conception of the good. Reciprocal reasons are simply not strong enough to contend with this question.

Enacting public policy on the basis of non-reciprocated reasons does not collapse the private realm. If certain values consistent with and entailed by the biological curriculum are embraced and taught publically, nothing is said on behalf of individuals, who remain free to believe what they like. But this free choice, rather than being consistent with and underdetermined by the values promulgated by the state, conflicts with publically-held values. For the person whose private beliefs conflict with the values disseminated publically, the education provided by the state is not a neutral repository of values or a means enabling him to choose his own conception of the good. In a liberal democratic regime, the content of public values is supposed to be given by what rational people desire regardless of their other interests or beliefs. This implies that anyone whose private beliefs conflict with public values are in some sense acting irrationally. On the contrary, we have already seen that no position concerning what ought to be taught in biology seems to be irrational—that is, in conflict with the values entailed by the just distribution of the primary social goods—in this special sense. Thus, if someone's beliefs were to conflict with the biological curriculum instituted by the state, his position might nevertheless be perfectly rational. Therefore, one must either give up Rawls's notion of legitimate

deliberation, or allow certain values that lie beyond the thin conception of the good to exist within the domain of public reason. In the next section, I will examine some responses of both kinds.

Aggregate Democracy and Non-Public Agreement: Viable Alternatives?

When a debate strays beyond the interpretation and extension of the thin conception of the good, resort must be made to non-public values. But where these values are not widely shared, deliberation is severely hampered. How can a person persuade another when he has no common grounds or values to which to appeal in making his case? When deliberation can no longer appeal to common or shared premises, resort must be made to pure political power.

Wherever democracies are genuine expressions of the will of the demos, as opposed to the will of corporate magnates, those with military force, or the rich, this political power is held by the majority of voters, who express their opinions by voting. Some have taken this procedure—a mechanism whereby the values enshrined in the public sphere via, law, educational policy, etc are produced by the aggregation of individual interests expressed in voting—to be a normative and descriptive model of democracy. Politics, on this view, is merely the collective expression of individual opinions through voting. Young calls this political model "Aggregate Democracy." Describing this view, Young says "The goal of democrat decision-making is to decide what leaders, rules, and policies will best correspond to the most widely and strongly held preferences" (Young 2000, p. 19). What other alternative is there, when appeals to a common good are not strong enough to resolve disputes concerning public values? Rather than attempting to conceptualize democracies as resting upon a set of values and concepts that express a genuine common good in accordance with public reason, perhaps it is more realistic to see democracy instead as a space in which individual interests are aggregated, and the goods embraced and taught by the state are beholden to those interests.

Young identifies a number of problems with this model of democracy. Early contractual conceptions of democracy were attempts to legitimize and institutionalize freedom as one of the motives of the establishment of the state. This conception relied on the idea that the worldview enshrined in the state could be limited to a set of basic and universal human interests. But if I am right in suggesting that public reason is insufficient in determining the right outcome to some political disputes, then something more than the articulation of basic rights and primary social goods is needed. The indeterminacy of the preferable values over and above the thin conception of the good is overcome, on the aggregate view, by the will and interest of the majority. The content of that more robust good that is expressed in the state is therefore merely the result of the competition of private interests in the political marketplace. It follows that the good expressed in the state is not the consequence of deliberation or reasoning, but of mere political interest. Those whose views get expressed, on this conception of democracy, are not and cannot be interested in finding some common ground with other citizens. But if the citizens of a regime are not united by common goods, and cannot appeal to reasons which all would find legitimate, what binds them together except the convenience of the political arrangement and fear of the retribution of the state? For Rawls, the legitimacy of a political outcome depends upon the public legitimacy of the reasons by which the legislators justify their political actions. They must reasonably think that the reasons given for their political actions are sufficient and acceptable to others as citizens. By contrast, aggregate democracy implies that there are no genuinely public spaces—that is, spaces in which people from a variety of backgrounds, ethnicities, and cultural affiliations can deliberate together about the values that ought to be embraced and disseminated publically.

I think that Young's notion of “Deliberative Democracy” is a promising alternative. According to Young, deliberative democracy “[i]s a form of practical reason” (Ibid, p. 22). This is what she says about it:

> Participants in the democratic process offer proposals for how best to solve problems or meet legitimate needs, and so on, and they present arguments through which they aim to persuade others to accept their proposals. Democratic process is primary a discussion of problems, conflicts, and claims of need or interest. Through dialogue others test and challenge these proposals and arguments… Participants arrive at a decision not by determining what preferences have greatest numerical support, but by determining which proposals the collective agrees are supported by the best reasons. (Ibid, p. 23)

Rawls's conception of liberal democratic procedure has a great deal in common with Young's notion of deliberative democracy. Rawls's attempt to articulate and demarcate a distinctly public reason is an attempt to give voice to the kinds of deliberative practices he takes to be appropriate to democracies. However, we have seen that Rawls's concept of public reason is not robust enough to arbitrate in key political debates concerning education. This difficulty is not unique to Rawls. All theorists of deliberative democracy have to wrestle with the legitimacy of the forms and modes of reasoning used in political debates. In efforts to overcome the impotence of thin notions of public reason, theorists have often emphasized the need for shared beliefs, goals and values as a necessary ground for effective deliberation.

For example, Miller has argued that nothing short of national pride can ground effective deliberation (Miller 1995, pp. 96-98). Mansbridge argues that participatory democracy is only appropriate where many kinds of goals, interests, and premises lying outside the primary social goods are widely shared (Mansbridge 1980, p. 19). Though one could criticize these specific formulations, the basic idea is hard to refute—without certain shared notions, values, or premises, deliberation looks completely useless. The integrity of an argument depends on two things: 1) the strength of the support which the premises lend the conclusion, and 2) the soundness of the premises. Without some agree-

ment upon both of these criterion, there is no way to convince someone of the truth of a conclusion to arguments of this kind.[10]

There is nevertheless a problem with this kind of theorizing. First, insofar as appeals to nationalism or other kinds of values or goals are supposed to prevent deliberative paralysis, they belong to the domain of non-public reason. Remember that the legitimacy of a political action, for Rawls, depends upon our belief that the "reasons we would offer for our political actions… are sufficient, and we also reasonably think that other citizens might also reasonably accept those reasons" (Rawls 1997, p. 578). A citizen is reasonable when he offers fair terms of political cooperation to his fellow citizens—when, in other words, he deliberates publically in such a way that the reasons for which he acts would in principle be legitimate to others, in spite of radically different comprehensive belief systems, values, or goals. Yet Mansbridge and Miller think that deliberative democracy can only persist on a shared basis of non-public beliefs and values. Because common ground of this kind cannot be guaranteed in a pluralistic society, deliberation on the basis of shared beliefs and values is likely to suppress and exclude the opinions of those who do not share those beliefs or values. One cannot merely assume a unity of values that transcend and render impotent differences of situation, culture, and identity. Rather than enabling pluralistic deliberation, a model which depends upon a broad non-public basis of agreement will only enforce the hegemony of dominant normative frameworks and alienate and marginalize groups which do not share the values of the majority.

A Modified Concept of Public Reason

I hope to have shown that neither aggregate models of democracy nor restricted notions of deliberation are desirable alternatives to Rawls's concept of public reason. On the other hand, Rawls's concept of public reason remains problematic. Engaging in public reason, for Rawls, requires citizens to abstract from their comprehensive systems of belief and ethnic or cultural affiliations. They

must justify their claims via reasons which anyone, qua citizen, would accept. However, we have seen that there are political disputes concerning which public reason—the use of reasons that satisfy the criterion of reciprocity—is insufficient. Rather than resign ourselves to the inevitable intrusion of irrational private values in the public sphere, however, this weakness in Rawls' concept spurs us to formulate a more inclusive and robust model of public reason.

For Rawls, public reason is restricted in scope to deliberation concerning the kinds of things that anyone, regardless of religious, ethnic, or cultural affiliation might accept. This means that public reason does not concern the ends or values for which people act, but only the distribution of the primary social goods, which function as all-purpose means. I propose that the subject of public reason be extended, in certain cases, to the values, ends, and claims which lie outside the scope of the thin conception of the good. There are instances in which political deliberation must include debates concerning the merits of various non-public ends. In Rawls's political theory such debates are essentially private and hence have no place in the public sphere. Rawls restricts public reason to deliberation about means. In doing so, he deliberately contrasts such reasoning with reason concerning ends. This is a basic tenant of liberalism—a view according to which the role of the state in an individual's life is restricted (ideally) to protecting certain rights and liberties, beyond which the individual is free to live the life he chooses.

The hope is that by restricting liberal deliberation to the means by which an individual chooses his lifestyle—equal rights and liberties, access to education, etc.—a clear private domain free from the interference of the state can be articulated and protected. If the merits of certain ends and values are discussed and evaluated publically and enshrined in policy, it seems that individual's private lives are unduly restricted. This would be a legitimate fear concerning the expanded concept of public reason if all values and ends lay open to it. However, the ends that are the legitimate subject of public reason are restricted to those that are consistent

with the basic tenants of liberalism. In other words, the rights and liberties of individuals and whatever other means might belong to the domain of public reason are given absolute priority over the ends towards which those rights and liberties might be put. If a value or end violates some person's basic rights, then it does not belong to the domain of public reason.

This is an important point which needs to be emphasized. The expansion of the concept of public reason does to some extent shift the line dividing the public and private realms. My modified concept of public reason does perhaps create some area of overlap between these two domains. However, this does not mean that no distinction between the public and private realms is possible. It does not imply that the whole private domain now lies open to the public. For example, the question of the relative merits of beer and wine is a question that clearly belongs to the domain of private preference. The equal distribution of rights, on the other hand, is clearly a public issue. The expansion of public reason applies more to cases concerning, for example, educational policy—issues in which multiple alternatives are consistent with the thin conception of the good but cannot both be implemented. In such cases, decisions must be made on the basis of the merits of the values or ends at stake in both.

With respect to Rawls' concept of public reason, this means that political justification will not be limited to reasons which a person reasonably thinks another person, as a free and equal citizen, might accept reasonably accept. Rather, deliberation must be thrown open to argument about the kinds of reasons to which citizens can legitimately appeal in political discussions.[11] Expanding political deliberation in this way will successfully bring questions like what ought to be taught in biology into the domain of public reason. Rather than being the outcome merely of aggregated political opinion, supporters of various practical proposals can display the values promoted in their respective educational policies, and can argue for the merits of those values. Furthermore, this enables and supports the inclusion of non-accepted forms of deliberation and kinds of reasons within the public realm. Rather than having

to justify a position from within a specific deliberative framework using a restricted set of accepted procedures and reasons, these individuals or groups have the opportunity to discuss the merits of alternative political positions by arguing for the merits of the premises and reasons which they think others ought to accept. In other words, they do not merely have to work within a majoritarian normative framework. They can challenge this framework while remaining within the domain of public reason.

Consider an example. Discussing her notion of public reason, Young mentions the production of narrative or situated knowledge as one deliberative technique. Briefly, situated knowledge is knowledge concerning the phenomenological experience of individuals or groups—a kind of knowledge unique to specific situations and therefore not universal. The legitimacy of the use of narrative or socially-situated knowledge in public deliberation has often been questioned.[12] In virtue of its non-universality, how can this kind of knowledge belong to the public domain? Though it is sometimes true that claims to phenomenological otherness have disrupted deliberation, it is also the case that phenomenological descriptions can strengthen deliberation and expand its purview. Rather than creating social barriers, drawing attention to social differences can enable participants to build bridges of communication. The practice of narrative can be used to demonstrate and persuade others of the legitimacy of appeals to phenomenological facts as reasons in public discourse. Much of the literature exploring the relationship between phenomenology and social criticism is engaged in exactly this kind of project.[13] Discussions of difference enable socially marginalized groups and individuals to challenge spurious appeals to the common good that often only serve to enforce the hegemony of a dominant social group and or economic class. The expansion of the concept of public reason enables these groups and individuals to challenge the dominant deliberative framework, which would exclude the sorts of reasons to which these groups might like to appeal, while remaining within the domain of public reason.

Conclusion

I think that this is a satisfactory answer to the problem of deliberation. Rawls' notion of public reason is not always sufficient for settling political questions. But this does not imply that deliberation ends at the boundaries of the thin conception of the good. We must expand the notion of deliberation in such a way that public reason is not limited merely to the thin conception of the good but nevertheless remains distinct from the private sphere. The best way to do this is to make the reasons native to the public sphere open to the deliberative process itself. Instead of only appealing to reasons that an individual sincerely thinks that another, as a citizen oriented to the advancement of public interest, might accept, individuals are free, within the public domain, to argue for the merits of certain kinds of reasons consistent with but not limited to the thin conception of the good. Their positions still have to be supported by good argument, as opposed to thoughtless appeals to private interests. This means that educational policy can be debated and decided on without paralysis arising from the limited scope of Rawls' notion of public reason. Furthermore, the expansion of public reason enables the political recognition and inclusion of groups where an unchallenged normative deliberative framework would often unwittingly marginalize their interests and appeals. The inclusion of discussion concerning the merits of various kinds of reasons within the public sphere means that reasons that lie outside this framework do not represent the illegitimate encroachment of private interest on the public sphere, but are instead central to the concept of public reason itself.

Notes

1. Locke and Mill are good examples.
2. The just distribution of the primary social goods is inconsistent with certain conceptions of the good. For example, a conception which denied some group or individual access to basic goods like rights would certainly conflict with a liberal democratic state. However, this is not inconsistent with the principles of liberalism, since such ideologies are not recognized as legitimate within a liberal framework.

3. To think of a citizen as a citizen is to think of him as the bearer of a set of interests related to the primary social goods and not as a person affiliated with religious or cultural associations or inhabiting other non-public social roles.
4. I call this kind of disagreement interpretive because the values and goods agreed upon are often inscribed in a constitution. Thus, this disagreement will sometimes concern the constitutionality of a given piece of legislation.
5. Rawls introduces his own conception of public reason via the Original Position and the Veil of Ignorance, but he claims that this is only one of many legitimate forms.
6. A reasonable reason is one which would be accepted by citizens offered fair and equal terms of social cooperation.
7. It does not have to be the case that the reason is, in fact, accepted. An unreasonable citizen might reject a reasonable reason. It merely has to be the case that the person appealing to the reason reasonably thinks that it is acceptable to others on the basis of their interest in the just distribution of the primary social goods.
8. Rawls adds a proviso which needs to be mentioned here: public reason only applies to situations concerning fundamental and important political values and interests. It is obvious that public reason cannot resolve questions about the beautification of the landscape of a government building, for example. This does not demonstrate the inadequacy of public reason, however, because public reason is not designed to apply to such questions.
9. Both YEC and evolution would have to be treated as equally legitimate theories, because this is the only way that neutrality with respect to non-public values could be maintained.
10. This depends, of course, on the assumption that the person to whom the argument is directed does not already believe the conclusion. This does not seem controversial.
11. The kinds of reasons whose merits are under consideration are limited to reasons consistent with the basic tenants of liberalism and democracy. Reasons like "I just don't like x" are not candidates for public deliberation because these systematically exclude those who like, are, or practice x. One can criticize x, but only in such a way that those who adhere to it have a chance to give counter-reasons and are therefore included in the deliberation.
12. Because socially-situated knowledge fails the criterion of reciprocity, it would at least be inconsistent with Rawls' conception of public reason.
13. See Sartre's Anti-Semite and Jew, Beauvoir's The Second Sex, and Fanon's Black Skin White Masks.

Bibliography

Mansbridge, Jane. (1980) *Beyond Adversary Democracy.* (New York: Basic Books)

Miller, David. (1995) *On Nationality.* (Oxford: Oxford University Press)

Rawls, John. (1968) "Distributive Justice: Some Addenda," in: S. Freeman (Ed), *John Rawls: Collected Papers*, pp. 154-175 (Cambridge: Harvard University Press)

Rawls, John. (1997) "The Idea of Public Reason Revisited," in: S. Freeman (Ed), *John Rawls: Collected Papers*, pp. 573-616 (Cambridge: Harvard University Press)

Rawls, John. (1999) *A Theory of Justice*. (Cambridge: Harvard Publishing)

Young, Iris Marion. (2000) *Inclusion and Democracy*. (Oxford: Oxford University Press)

No Retrosuicide Paradox in Time Travel

Walter Orozco

Introduction

Thoughts of time travel seem to capture people's imagination; plenty of movies and books have been produced surrounding the subject. Perhaps in the future time travel to the past will be an actuality. However, long before time travel can be an actuality, it needs to be a possibility. Philosophical discussion is suited for thoughts of the possibility of time travel.

It is a common occurrence that thoughts of time travel quickly turn into thoughts about killing someone. I am of course *half* joking. Philosophical consideration of time travel results in interesting puzzles of how someone could go to the past and manage to kill a younger version of himself. Such an act of killing a younger version of oneself is what Peter Vranas calls *retrosuicide* (Vranas 2009, p. 520). The puzzle (or paradox) comes from realizing that killing a younger version of yourself would mean that you could not grow up to be the time traveler that tries to kill himself in the first place; loosely this is the *retrosuicide paradox* (we can get into the precise formulation later on) (Vranas 2009, p. 520).

David Lewis considered a similar paradox, the *grandfather paradox* (Lewis 1976, p. 149). It runs along similar lines, but rather than attempting retrosuicide the time traveler tries to kill his grandfather (Lewis 1976, pp. 149-151). The retrosuicide paradox and the grandfather paradox are not the same, but Vranas's paradox was inspired by Lewis's paradox (Vranas 2009, p. 520). However, Vranas does not think that the solution (so called "standard solution") that works for the grandfather paradox will work for the retrosuicide paradox (Vranas 2009, p. 520-521). The *standard*

solution "…explains the paradox away as due to an equivocation [about what the time traveler can and cannot do]…" (Vranas 2009, p. 520). Getting into the details of the standard solution now would be too cumbersome. For now let it suffice to say that Vranas thinks his retrosuicide paradox (henceforth RP) cannot be solved by the standard solution. Ultimately, I disagree with him.

More precisely, I hold that Vranas is unfair to the standard solution as applied to the RP and that a standard-like solution can be applied to solve the RP. The first order of business will be to make clear general terminology and conditions of the RP; this will be done in section 1. In section 2, I will carefully lay out the RP. Section 3 will be dedicated to understanding a key concept in the standard solution, namely compossibility. Section 4 will present the standard solution as applied to the RP according to Vranas. In section 5 I will explain how Vranas is unfair to the standard solution. In section 6 I will show how a standard-like solution can be used to solve the RP. I will use section 7 to answer the concern that for the standard-like solution to work a particular metaphysics of time must be assumed to be true. Finally, in section 8 I will address the concern that what can be said of the RP can also be said of the grandfather paradox. Let us move to our first order of business.

1. General Terms and Conditions

In a general way a *paradox* can be defined as, "an apparently sound argument with an unacceptable conclusion (cf. Sainsbury 1995, p.1 as cited by Vranas 2009, p. 522). For our discussion time travel will solely refer to time travel into the past; and, when talking about possibility we will be concerned with physical possibility (Vranas 2009, p. 522). Physical possibility has to do with what is possible given the physical laws of the actual world. Talk of possibility will sometimes turn into talk about possible worlds; this kind of talk is just a useful tool to consider possibilities. Possible worlds considered in this discussion will have to be ones with the same physical laws as the actual world.

Vranas believes that concluding that time travel is impossible is an unacceptable conclusion, "…because time travel is compatible with General Relativity" (Vranas 2009, p. 522).[1] I take it that this compatibility leads Vranas to believe that time travel is physically possible. For the sake of this discussion, I think it best if we simply assume that time travel is physically possible; the possibility of time travel is not in contention in this paper.

Physical possibility will also preclude talk of resurrection (Vranas 2009, p. 524). Thus, talk of death will be such that death is, "permanent and irreversible" (Vranas 2009, p. 530). Vranas reasons that, "…resurrection, being a miracle, does seem to be impossible…" (Vranas 2009, p. 524). This stipulation precludes a "short-cut solution" that would appeal to resurrection. Such a solution would seem to miss the spirit of the paradox. Loosely speaking, when I refer to "the spirit of the paradox" I am referring the intuitive inclination we seem to have about what the paradox is saying. With these terms and conditions made clear it's time to layout the RP.

2. The Paradox

In this section I will first present the RP and second I will clarify unclear terms and explain how the RP does seem to be a paradox.

First, Vranas lays out the RP in the following way:

(P1) If time travel is possible, then it is possible for me to be in a retrosuicide-propitious situation.

(P2) Necessarily, if I am in a retrosuicide-propitious situation, then in that situation I can kill my younger self.

(P3) Necessarily, if I am in a retrosuicide-propitious situation, then in that situation I cannot kill my younger self.

Thus:

(C) Time travel is impossible. (Vranas 2009, p. 521-522)

Two terms need to be made clear here: First, a *retrosuicide-*

propitious situation is a situation in which it *seems* that your odds of killing your younger self are extraordinarily good (Vranas 2009, p. 521). Second, *younger self* refers to an earlier temporal stage of a person, in this case, the time traveler (Vranas, 2009, p. 521).

At face value it seems that the RP is truly a paradox. We can understand it as a *reductio ad absurdum*: if our initial assumption leads to a contradiction then that assumption must be false. Here our initial assumption is that "time travel is possible" from P1. That leads to P2 and P3, which seem to be contradictory; if both are true then we are forced to say that the time traveler in a retrosuicide-propitious situation can and cannot kill his younger self. Thus, we must conclude that time travel is not possible. That conclusion is unacceptable given that both sides, Vranas and I, believe time travel is possible. Then, by definition, the RP is indeed a paradox. Since we have a paradox on our hands, the next step is to explain the notion of compossibility so that later we can see why Vranas believes that the standard solution fails in solving the RP.

3. Comprehending Compossibility

The standard solution makes use of *compossibility* to show that there can be equivocation when talking about what the time traveler can or cannot do. By clearing up the equivocation the standard solution permits us to say that the time traveler both *can* and *cannot* commit retrosuicide without committing us to a contradiction. So to understand the standard solution we need to understand compossibility.

Compossibility can be understood as something that is *possible* in combination with something else; that 'something else' is the relevant facts. Lewis puts it in the following way, "To say that something can happen means that its happening is compossible with certain facts" (Lewis 1976, p. 150). Which facts are determined by context: "What I can do, relative to one set of facts, I cannot do, relative to another, more inclusive, set" (Lewis 1976, p. 150).

For example, to say that event *alpha* can happen is to say that it is compossible with the set of facts {a, b, c}. While to say that event *alpha* cannot happen is to say that it is not compossible to a more inclusive set of facts, say the set of facts {a, b, c, d, e}. Notice that the more facts that are considered the less the event seems to be possible.

Lewis gives us a more concrete example of how to understand compossibility (Lewis 1976, p. 150). Imagine someone were to come to me and ask whether I can speak Finnish. I could, *honestly*, say that I can speak Finnish. Indeed I can, relative to the facts of my vocal cords, larynx, nervous system and other facts of my anatomy. In other words, my speaking Finnish is compossible with the facts of my vocal cords, larynx, nervous system and other facts of my anatomy (Lewis 1976, p. 150).

If that same person were to ask me how to say "I love you" in Finnish, I would have to admit that I do not know how to speak it. I could, *honestly*, say that I *cannot* speak Finnish. Indeed I cannot, given a more inclusive set of facts than before, namely that I have no training in the Finnish language. That is to say, my speaking Finnish is *not* compossible with the more inclusive set of facts, which not only include facts about my anatomy but also facts about my lack of training (Lewis 1976, p. 150). So I can honestly—or rather, truly—say that I can and cannot speak Finnish without committing a contradiction.

Someone might demand to know, "Which is it then? Can you or can't you?" If you want to look at the broadest set of facts, then I *cannot* speak Finnish. But imagine how odd it would be if someone were to offer me lessons in Finnish and I turned her down because I say that I cannot speak Finnish. I would hope the instructor would say, "Of course you *can*! I can teach you." Notice then that context—determined by the relevant facts of the situation—matters when answering questions of ability (Lewis 1976, p. 150). Compossibility is key to the standard solution and to why Vranas does not think it works for the RP, as we shall see in the next section.

4. Standard Solution a la Vranas

In this section I will first lay out the standard solution the way Vranas sees it and then I will next explain why Vranas thinks it fails.

Our initial assumption is:

(P1) If time travel is possible, then it is possible for me to be in a retrosuicide-propitious situation. (Vranas 2009 p. 521)

In order to apply the standard solution to P2 and P3, we need to point out the relevant facts in P2 and P3 and explain how "can" becomes equivocated between the two premises. In attempting to apply the standard solution we get P2' and P3':

(P2') Necessarily, if I am in a retrosuicide-propitious situation, then in that situation I can kill my younger self, in the sense that the relevant features [i.e. relevant fact] of the situation—namely (for example) that he is asleep [my gun is fully functional and I am at close range] etc.—are compossible with my killing [my younger self].

(P3') Necessarily, if I am in a retrosuicide-propitious situation, then in that situation I cannot kill my younger self, in the sense that the relevant features of the situation—namely that *I don't kill him*, that *he is an earlier stage of mine*, and (for example) that he is asleep etc.—are not compossible with my killing [my younger self]. (Vranas 2009 p. 522-523)

Now it should seem that no contradiction is made when saying that the time traveler can and cannot kill his younger self. Given a less inclusive set of facts the time traveler can kill his younger self and given a more inclusive set of facts he cannot kill his younger self. But Vranas brings in a concern, "But what if it is necessarily *impossible* for me to kill my younger self? Then my killing him is necessarily *not compossible* with anything…"

(Vranas 2009, p. 523). Vranas points out that "impossibility entails inability;" and, surely if something is impossible, it *cannot* be done (Vranas 2009, p. 531). So if something cannot be done at all, then it is not compossible with anything.

If this line of reasoning is correct, then the consequent of P2' is negated, because it seems that no set of facts can be compossible with the claim that the time traveler can kill his younger self. By *modus tollens* (henceforth *M.T.*) it follows that a time traveler cannot be in a retrosuicide-propitious situation. If it is impossible to be in a retrosuicide-propitious situation then via *M.T.* on P1 it follows that time travel is not possible. That's an unacceptable conclusion, thus the paradox is restored.

In order to prevent this chain of *M.T.* inferences Vranas concludes that the standard solution relies on a *hidden assumption*, "…that it is *not* necessarily impossible for [the time traveler] to kill [his] younger self…" (Vranas 2009, p. 523). And that the standard solution makes that assumption without argument, so the hidden assumption could be false and as such would cause the standard solution to fail. However, I feel that this is a misunderstanding of how the standard solution works.

5. Delineation Termination

I this section I want to first make it intuitively clear how the standard solution works on a paradox similar to the RP, but how it ultimately cannot work on the RP. Second, I want to explain how Vranas's attempt to apply the standard solution to the RP is unfair.

To achieve my first task in this section I will present a small hypothetical, let us call it H1. Keep in mind that even in this hypothetical we are still talking about physical possibility and we are holding resurrection as impossible. Picture H1: The room is dark, no one is home, the victim has been stalked carefully, and your gun is loaded and fully functional. The victim is helplessly asleep in front of you. As moonlight gently floods into the room you recognize the victim. You carefully aim and fire.[2]

Given only the delineation of facts in H1, it should be clear

that you can indeed kill your victim; in fact, it is flat-out possible. Nothing is stopping you, everything is in your favor, and you have everything it takes. So at this point no "inconsistency alarms" should be going off in your head.

However, now imagine H1+. H1+ has all the facts in H1 but also has one more interesting fact. The person you recognized is you, as a child. Suddenly it becomes evident that you cannot kill the victim, otherwise you would have never grown up in the first place to attempt to commit retrosuicide. Given the delineation of facts in H1+ it is impossible for you to kill the victim.

Therefore, it is true that you can and you cannot kill your victim, "but under different delineations of the relevant facts. You can reasonably choose the narrow delineation [H1] and say that [you] can; or the wider delineation [H1+] and say that [you] can't" (Lewis 1976, p. 151). And all the while a contradiction is being avoided.

I am using these two hypotheticals to achieve two things. First, I am attempting to recreate the standard solution, as it would work in a paradox similar to the RP. Second, I want to show that the more narrow the delineation of fact, the wider the possibility in question becomes (e.g. H1). However, the wider delineation of facts, the more narrow the possibility in question becomes (e.g. H1+).

These latter points explain why the standard solution does not work on the RP. From the very start, the RP tells us that the victim is a younger self of the retrosuicidal time traveler. In other words, the RP forces us to look at the broadest delineation of facts and thus consider a very narrow possibility. Facts presented in P2 of the RP are like those present in H1+ and not like those facts present in H1. Thus, it seems unfair to even attempt to apply the standard solution to the RP as it formulated by Vranas. The standard solution would only work if there were different delineations of facts between P2 and P3. However, the standard solution does not makes a hidden assumption, rather the standard solution cannot even get off the ground given Vranas's formulation of the RP. Furthermore, it would be odd if the standard solution were to

make the hidden assumption, that retrosuicide is possible (Vranas, 2009, p. 521), because when it is applied to the RP it concludes that retrosuicide—given a very inclusive delineation of facts—is actually impossible.

One might be tempted to say that this becomes a problem of formulation (Vranas, 2009, p. 524). Imagine if we were to reformulate the RP so P2 did not already include facts about the time traveler's *younger self*, but only addressed "the boy in front of the time traveler". Then the standard solution could be applied because there would be different delineations of facts between P2 and P3. But reformulation is not the answer, because to reformulate the RP would be to address a similar, yet different paradox. However, the RP as it is actually formulated by Vranas, cannot directly be solved by the standard solution.

Even if the RP cannot be directly solved by the standard solution we can still use Lewis's claim that it is impossible to change the past to solve the RP (this is what I'm calling the standard-like solution) (Lewis 1976, p. 150).[3]

6. Don't Ever Change

My argumentative strategy here will be to solve the RP by showing that P2 is false; thus, RP would fail to follow. It seems reasonable that since Vranas's RP was inspired by Lewis's formulation of the grandfather paradox (Vranas 2009, p. 520), that Vranas has a similar understanding of time. As such, I will take that for granted, at least for now.

Remember that P2 is, "Necessarily, if I am in a retrosuicide-propitious situation, then in that situation I can kill my younger self" (Vranas 2009, p. 521). In a retrosuicide-propitious situation the odds are incredibly good that retrosuicide can happen, but it is not guaranteed. In fact, if Lewis is right, and time cannot be changed, then no matter how good the time traveler's odds of committing retrosuicide are, we already know that he will fail (Lewis 1976, p. 150). P2 forces us to look at the broadest set of facts, not only the retrosuicide-propitiousness—which includes

the time traveler's proximity to his victim and his fully functional gun—but also the fact that the victim is the younger stage of the time traveler, and that victim grows up to be the time traveler. Given these facts we are forced to conclude that the time traveler did not kill his younger self. He failed because it is impossible to change the past. Paraphrasing Lewis: the events of the past timelessly do not include the time traveler committing retrosuicide (Lewis 1976, p. 150). Indeed, given the set of facts present in P2, it is impossible for the time traveler to kill his younger self. And as Vranas reminds us, impossibility entails inability, and as such P2 is false. Now the RP is dispelled because it ceases to be an apparently sound argument.

Two difficulties arise in arguing that P2 is false. The first is that it might seem that I made an unfounded move in resolving the RP by asserting that Vranas understands time in the same way as Lewis. The second is that what can be said of the RP can also be said of the of the grandfather paradox.[4] In the following sections I will address each concern in their respective order.

7. Committing to Eternalism

The way I resolved the RP in the last section was to appeal to Lewis's understanding of time and claim that since Lewis inspired Vranas in his formulation of the RP (Vranas 2009, p. 520), then Vranas could have the same view of time travel as Lewis. And someone would be right to point out that that's no better than simply assuming that Vranas views time the way Lewis does. In order to answer to that concern I want to attempt to show that by the very nature of the RP, Vranas commits himself to a metaphysics that would not permit for changes to the past. That metaphysical view is *eternalism*.

Eternalism holds that time, "… is a fourth dimension essentially constitutive of reality together with space. All times, past, present and future, are actual times just like all points distributed in space are actual points in space" (Hunter). In other words, time is just another dimension in addition to the three dimensions of

space, and past, present and future all enjoy the same kind of reality. This sounds exactly like what Lewis says, "[t]he world… is a four dimensional manifold of events. Time is one dimension of the four…" (Lewis 1976, p. 145).

It seems then that Lewis is an eternalist. If I can commit Vranas to eternalism then I have a stronger case that Vranas understands time the way Lewis does. If I can manage that, then my solution of RP by the immutability of time goes through.

There are two main reasons why I argue Vranas is committed to eternalism: one, the RP deals with time travel; and, two, his definition of "younger self". I will briefly explore each of these reasons.

First, time travel is typically most compatible with eternalism; because in eternalism the past is as real as the present; thus, it gives the time traveler a "somewhen" to go to.

Second Vranas defines "younger self" as an earlier temporal stage of the time traveler and notes that, "Three-dimensionalist may be unhappy with my talk of stages, but Effingham and Robson (2007) and Sider (2001, pp. 101-9) have argued that three-dimensionalism is incompatible with time travel" (Vranas 2009, p. 529). Three-dimensionalism claims that, "… objects are three-dimensional, wholly present at each moment" (Hawley 2004); in other words, it claims that what makes up the whole of an object is completely included in the present. So three-dimensionalism does not allow objects (including humans) to have temporal past parts. In direct contrast, four-dimensionalism is the view that, "… objects are extended in time and are thus four-dimensional, having different [temporal stages] at different times" (Beebee, Effingham and Goff 2011, p. 79). So it would seem that Vranas requires four-dimensionalism to make his definition of "younger self" work. And four-dimensionalism fits most comfortably into eternalism, which views time as four-dimensional.

Neither of these reasons are knockdown reasons to confirm Vranas's view on time. Since Vranas is not explicit about his view these reasons seem to provide evidence that Vranas is committed to eternalism. As such the solution in section 6 goes through.

8. Different Paradoxes

Another concern that arises is what can be said of the RP can also be said of the grandfather paradox; thus, I could have inadvertently made the standard solution fail when applied to the grandfather paradox. It would be odd for me to say that, since even Vranas claims that, "…what Lewis says about the *grandfather* paradox is fine" (Vranas 2009, p. 530). Furthermore, if the standard solution fails when applied to the grandfather paradox then there would seem to be no point in defending anything like the standard solution in the first place, and thus my project would be rather pointless. So in this section I want to explain how this concern arises and then attempt to resolve it.

For the sake of explaining the concern at hand I will try to formulate a formal representation of the grandfather paradox according to what Lewis says (Lewis 1976, pp. 149-151).

(GP1) If time travel is possible, then GP2 and GP3 are both possible.

(GP2) Tim [the time traveler] can kill Grandfather

(GP3) Tim cannot kill Grandfather

Thus:

(C) Time travel is impossible (Lewis 1976, p. 149-151).

Remember that the way I argued against the RP was to say that P2 of the RP smuggled in facts about the time traveler and his younger self and thus the delineation of facts in P2 were so inclusive they made P2 false. However, if we look at GP2 we see that there is a fact about the time traveler and his grandfather. So by similar logic used to solve the RP, it would seem that I am forced to say that GP2 is false. And as such, there is no sense in which the time traveler *can* kill his grandfather. Thus, the standard solution either fails in the grandfather paradox or is not needed to solve the grandfather paradox. This is the concern I need to answer.

It seems that GP2 also manages to sneak in facts in the same way that P2 of the RP does. But there are particular distinctions

between the two paradoxes that make this not necessarily true. First, the type of possibility being considered in the RP and the type of possibility being considered in the grandfather paradox are different. The RP is concerned with physical possibility, which includes the impossibility of resurrection (Vranas 2009, p. 522-524). On the other hand, the grandfather paradox has to do with a broader metaphysical possibility; as Lewis puts it, "a possible world where time travel took place would be a most strange world, different in fundamental ways from the world we think is ours" (Lewis 1976, p. 145). So it could be that when considering the grandfather paradox we could consider a possible world in which resurrection is possible so that the time traveler can kill his grandfather, and the grandfather comes back to life and begets the time traveler's father. And thus we have a way in which GP2 is true. However, this type of solution seems to miss the spirit of the grandfather paradox.

So if we hold resurrection as a fixed fact in the grandfather paradox, does GP2 still turn out false? Not necessarily. It seems to me that the term "younger self" sneaks in more facts than the term "grandfather." Surely killing one's grandfather is possible without considering time travel, but killing a younger version of oneself is not possible without considering some sort of time travel. So the term "younger self" in P2 of the RP forces us to consider the time traveler's ability in relation to a more inclusive delineation of facts, while the term "grandfather" in GP2 need not force us to consider any broader delineation of facts than those relevant to the time traveler's ability. Thus, in the grandfather paradox we can say the second premise is true given a less inclusive set of relevant facts; which is something that cannot be done in P2. It is in this manner that the grandfather paradox differs from the RP, and as such what is said of the RP need not also be said of the grandfather paradox.

Conclusion

Like I said at the very beginning, thoughts of time travel seem to

turn into thoughts about killing someone; for example, someone's younger self or someone's grandfather. Vranas presents us with this case of retrosuicide and the apparent paradox. What makes the RP even more interesting is that it cannot be solved in the "standard" way of appealing to equivocation of the word "can". But rather is solved by demonstrating that since the past cannot change, the RP fails to be a sound argument. The RP is another interesting puzzle of time travel, which might simply seem esoteric, but for us who dream about time travel the RP is another interesting puzzle to consider.

Notes

1. Vranas gives a long list of sources to backup the claim that time travel is compatible with General Relativity. The relevance of these do not seem as important here since we will simply be assuming that time travel is physically possible.
2. This setup was partly inspired by Vranas's way of defining retrosuicide propitious.
3. Vranas alludes to this solution but has no response in published work (Vranas 2009, p. 529).
4. Special thanks to Mark Balaguer for making this particular concern salient.

Bibliography

Beebee H., Effingham N. and Goff P. (2011). *Metaphysics The Key Concepts* (London and New York: Routledge)

Effingham, N., & Robson, J. (2007). "A Mereological Challenge to Endurantism," *Australasian Journal of Philosophy* 85, pp. 633-640.

Hawley, K. (2004). "Temporal Parts". Retrieved March 5 2015, from http://plato.stanford.edu/entries/temporal-parts/

Hunter, J. "Time Travel [Internet Encyclopedia of Philosophy]." Retrieved March 12 2015, from http://www.iep.utm.edu/timetrav/

Lewis, D. K. (1976). "The Paradoxes of Time Travel," *American Philosophical Quarterly* 13, pp. 145-152.

Sider, T. (2001). Four-dimensionalism: An Ontology of Persistence and Time (Oxford: Clarendon Press)

Vranas, P. B. M. (2009) "Can I Kill My Younger Self? Time Travel and the Retrosuicide Paradox," *Pacific Philosophical Quarterly* 90, pp. 520-534.

The Nonexistence of Things

Rena Goldstein

It is controversial to say that some objects do not exist. We generally suppose that mythical creatures do not exist, nor do characters in fairy tales or those from stories. James T. Kirk may seem like one such object, as does Titania, Queen of the Fairies, from *A Midsummer Night's Dream*, and so does the infamous detective Sherlock Holmes seem like another such nonexistent (or for our purposes, fictional) object. But as will be explicated, this is deeply problematic; for something to not exist seemingly means that there *are* some objects that do not exist. Sentences containing the name of a fictional object do not refer to a material object in the *actual* world.[1] 'Kirk' does not refer to a living human being or a physical object. In contrast, a term like 'Earth' has reference. 'Earth' refers to the round sphere orbiting the sun. Thus, we can evaluate the truth condition of the following sentence: "The Earth is square." This sentence is false. Likewise, the sentence, "The Earth is 70% water" is true. These two sentences are in part evaluated based on the reference of the object. We can determine the truth conditions of these sentences by observing that the Earth is in fact not square and made 70% of water. However, when names do not refer, evaluating the truth conditions of sentences containing them is difficult. How is it determined whether a sentence is true or false when it contains a name that does not refer to a material object? On what basis is it judged?

Alexius Meinong, an Austrian philosopher writing at the turn of the twentieth century, famously used two examples—the round square and the golden mountain—to describe possible and impossible objects. A possible object is an object that does not exist in the actual world, but has the possibility of existing. A classic example of a possible object is a golden mountain. There obviously does not exist a mountain made of gold in the actual

world, yet it seems possible for such an object to exist. It seems possible because the concepts of 'gold' and 'mountain' are not contradictory. An impossible object is just that—a contradiction. An impossible object is an object that cannot possibly exist. A classic example is a round square. Round squares are not possible, yet it seems that round squares have *some* form of substance. G.E. Moore put it like this:

> It seems as if purely imaginary things, even though they be absolutely contradictory like a round square, must still have some kind of being—must still be in a sense—simply because we can think and talk about them.... And now in saying that there is no such thing as a round square, I seem to imply that there is such a thing. It seems as if there must be such a thing, merely in order that it may have the property of not-being. It seems, therefore, that to say of anything whatever that we can mention that it absolutely is not, were to contradict ourselves: as if everything we can mention must be, must have some kind of being. (Moore 1953, as cited in Salmon 1998, p. 277)

A similar problem is thus extended to fictional objects: when someone utters a sentence containing a fictional name (which does not refer a material object), it seems like the sentence is referring to something that *is*. But what kind of being is a fictional object? I will argue that fictional objects exist as abstract objects (non-mental, non-physical objects), and that these abstract objects are mind-dependent.[2]

Meinong proposed a radical theory: some objects have a lower ontological class of existence. This idea has all but been rejected. It just doesn't make sense to talk about two objects, one that fully exists and one that only *somewhat* exists. Quine argues that Meinong obfuscates the word 'exist.' He writes: "Despite his espousal of unactualized possibles, he limits the word 'existence' to actuality…. [He] grants us the nonexistence of Pegasus and then, contrary to what *we* meant by nonexistence of Pegasus, insists that Pegasus *is*" (Quine 1953, p. 3). Meinong accepts the

notion of possible worlds, of which the actual world is merely one instance. Meinong also admits that Pegasus does not exist—i.e., Pegasus is not a part of the actual world, but resides in at least one other possible world—but, despite the fact that Pegasus does not exist, Pegasus still *is* in some sense of the word that does not imply existence. While some endorse Meinong's special classification of possible objects, philosophers like Quine and Lewis have shown that such a theory is not without issue.

Graham Priest's view is an improvement on Meinongism. His argument developed from discussions with Richard Sylvan, who coined the term 'Noneism'. Sylvan defended a version of Meinongism in his book *Exploring Meinong's Jungle and Beyond: An Investigation of Noneism and the Theory of Items* (*Exploring Meinong's Jungle* for short). Lewis critiqued Sylvan's *Jungle* in a paper entitled *Noneism or Allism?* He explained three noteworthy schools in the nonexistence debate: 'Noneism', 'Allism', and 'Some-but-only-someism'. According to Lewis, Noneists hold that "none of the controversial objects exist," where controversial objects include abstract objects and fictional entities, but exclude common objects like tables and chairs (Lewis 1990, p. 23). An Allist argues that *all* of the controversial entities exist. Finally, the Some-but-only-someists argue that only *some* of the controversial entities exist.

This paper will only focus on fictional objects like literary characters. I will not touch on problems dealing with numbers, imcompossible things, or impossible things. I do not suppose that the argument I bring forth will solve these other debates; the argument here pertains only to a conversation regarding fictional characters.

I will argue that Priest's view of existential predication is incomprehensible. In section I, I will explain Priest's version of Noneism. In Section II, I will look at two objections to Priest's view from Quine and Lewis, and give Priest's response. In section III, I will argue that Kripke's view is the right way to think about fictional objects and offer an improvement on his view by arguing that fictional objects are mind-dependent.

One might wonder whether the study of nonexistent objects is important. It is important to investigate what it means to make claims about fictional objects because it seems as though fictional objects have properties. It is common to utter statements like, "James T. Kirk captains the Enterprise," or "Sherlock Holmes is a detective," but on deeper exploration, as will be undertaken, making such claims entails certain ontological commitments.

Section I: Priest's Version of Noneism

According to Priest's version of Noneism, fictional objects are neither physical, mental, nor abstract (non-mental, non-physical things). They merely do not exist. Thus, James T. Kirk just *is* not. Priest writes, "For the noneist, indeed, to exist and to be are exactly the same thing. Holmes does not exist; Holmes is not. There exists/is nothing that is Sherlock Holmes" (Priest 2005, p. 108). Unlike Meinong, Priest asserts that fictional objects do not have a lower ontological status (or a lower status of being). They have no being whatsoever (Priest 2005, p. 14). The only things that exist are concrete objects, such as tables, planes, trees, and people. Sylvan held a stronger view of Noneism than that of Priest. He believed only present things exist; concrete past and future objects have the same existential status as abstract objects—none. Priest is clear when he asserts that he does not endorse this stronger view of Meinongism.

All other objects, such as: abstract objects, worlds, merely possible objects, impossible objects, incompossible objects and the like do not exist. And yet these non-existent, fictional objects can be (and are) characterized as having certain properties, like 'captains a Star Ship,' 'is a doctor,' 'is a detective,' 'travels in space,' 'lives at 221 B Baker St.,' etc. Object x has the properties it is characterized as having not at this world, α (the actual world), but at a world, w, that realizes the way the story is represented. Thus, Kirk has the properties he is characterized as having not at α but at a world, w, which realizes the way in which the story is represented when watching *Star Trek*. At w Kirk *is*—at w he is

characterized as existing.

Existence is expressed via an 'existence' predicate. Much like an object can satisfy the predicate 'is blue,' so too an object can satisfy the predicate 'is existing'. This entails that an object need not exist for it to contain properties. The ways in which existential and universal quantifiers are commonly understood are therefore no longer appropriate within these semantics. Usually the universal quantifier, "∀x," is read as "all x are such that"; and "∃x" is usually read as "there exists an x such that" or "there is an x such that." But for Priest's semantics, reading "∃x" as "there exists an x such that" is both imprecise and inaccurate. Priest claims that the existential quantifier should be understood in the following way: "∃x" should be read as "some x is such that." So ∃x(Px ∨ Qx) reads as: 'some x is such that either x is a P or x is a Q'. Therefore, to say that an object *w* does not exist in a world *does not* imply that *w* has a lesser grade of existence (as was the Meinongian view before Sylvan). He writes, "If [*w*] does not exist (at a world), then *w* simply *is* not (there)" (Priest 2011, p. 110). Existential commitment must be given explicitly by way of existential predication.

Therefore, an object's characterization as having certain properties is irrelevant to existential status. He writes, "Meinong insisted that the Sein (being) of an object is independent of its Sosein (properties)" (Priest 2005, p. vii). Thus, objects are specified by certain sets of conditions—i.e. 'captains a Star Ship,' 'lives in the 22nd century,' 'is brave,' etc. Priest asserts that the conjunction of these conditions can be written as A(x). The object, *k*, characterized as having these properties, has its characterizing properties, A(*k*), and whatever other properties might follow (Priest 2005, p. vii). This is the characterization principle (CP). The CP explains how one can know what properties a nonexistent object is characterized as having: "we know that objects characterized in certain ways have those properties, precisely because they are characterized in that way" (Priest 2005, p. viii). Characterization of nonexistent objects is expressed through an account of descriptions. From Priest, an object x with 'such and such properties' are

noun-phrases that refer to objects (Priest 2005, p. 82). Truth conditions are determined by an object's characterization. "Let A(x) be any condition; this characterizes an object, c_A. and $A(c_A)$ is true—maybe not at this world, but at other worlds. Which? Cognitive agents represent the world to themselves in certain ways" (Priest 2005, p. 84). Hence, $A(c_A)$ is true if and only if something satisfies the condition of A(x). In other words, an object characterized at world *w* is true if and only if the world is represented as such by the utterer. The representation may be an incomplete or inaccurate representation of this world, but, according to Priest, it is possible that such a world is just like that. What does an incomplete representation look like? I might imagine the *Enterprise* and represent the situation according to our world, in which the *Enterprise* certainly does not exist. And, also imaginably, there is a world wherein the *Enterprise* does exist in 2015. It is possible. In any case, Priest writes that "the representation is incomplete with respect to many details," (p. 84) such as the building of the *Enterprise* (perhaps 2015, perhaps never). In the actual world, the *Enterprise* has not been built; but there are worlds that realize the situation wherein the *Enterprise* is built in 2015. Thus representation is incomplete.

Priest realized a problem with the CP; namely that, when given full generality, anything we imagine can be brought into being. This is problematic. An object need not exist for it to be characterized as having certain properties. Entailed within the Noneist doctrine (what the Noneist is committed to, anyway) is that 'existence' is a predicate. It seems that one can simply characterize an object with the predicate of 'existence' and have it be so—suddenly, a new object exists at a possible world! Just like I might characterize an object, g, as 'wearing green' or 'was laughing' it seems I can just as easily characterize it as 'existing.' *G* is now phenomenologically present to me. It appears applying existential commitment is rather arbitrary. I can confer an existential commitment on any object I dream up in my imagination. I could apply 'is blue' and 'is existing' to a fictional object I just made up, and which I shall dub 'the digger Tralazar'—and now

the digger Tralazar is phenomenologically present to me.

It seems then that the CP must be restricted to certain contexts. The problem is to which class of contexts should be given priority and why? Priest provides a solution to the problem of the CP, namely that "the CP can hold unrestrictedly, provided only that its instances may hold, not at this world, but at others" (Priest 2005, p. viii). Thus, 'the digger Tralazar' contains the predicate 'is blue' and 'is existing' not in the actual world, but at a world that realizes the situation. Likewise, James T. Kirk may be characterized as 'existing' and 'captains a Star Ship' in another possible world. Priest asserts, "we just do not assume that an object characterized in a certain way has its characterizing properties at the actual world, only at the worlds which realize the way the agent represents things to be in the case at hand" (Priest 2005, p. 85). Objects are thus referenced and characterized through an act of intentionality. In other words, objects are characterized via the intention of the utterer. The statement 'Kirk exists' is true at a world that realizes the situation.

Exploring how Priest solved the problem of the CP in any greater detail is not relevant here. What is relevant is what is concluded from the CP. Fictional objects are characterized as having certain properties in some possible world, namely in those worlds that realize the situation about the object. According to Priest, this is how fictional objects are realized. In the actual world, there is no guarantee that the object in question really has these properties; but in some worlds, the object does contain these properties. Say we characterize an object as having the property of commanding a Star Ship. In the actual world, there is no object that has this property (for Star Ships don't exist). But in some possible world that realizes the situation (e.g. the world described in Roddenberry's script), the object does contain the property of commanding a Star Ship. Moreover, fictional objects may have other properties than they are explicitly characterized as having (Priest 2011, p. 116). While Kirk may be characterized as compassionate, there is a possible world in which he is ruthless, and a possible world in which he is wimpy, for example—for as

discussed earlier, representation is incomplete.

Priest endorses constant-domain semantics. Constant-domain semantics entails that objects are not dependent upon the world of which they are a part. There are three types of worlds: the actual, the possible, and the impossible. Intensional states directed toward impossibilities necessitate impossible worlds.[3] Take the concept of the round-square—but, of course, such an object is impossible. Impossible worlds are just those worlds that realize the content of these states (Priest 2011, p. 109). Likewise, there are also incomplete and inconsistent worlds.[4] And possible worlds merely realize objects that are possible, such as Captain Kirk. According to constant-domain semantics, the status of fictional objects does not affect the domain of objects. Plainly spoken, if I completely destroy an object in this world, it does not mean that the object no longer is in the domain of all objects. For example, suppose there is a decorative vase sitting on a table. My cat, Anima, jumps on the table and, in the process, knocks the vase off the table. The vase is now smashed to bits. Under constant-domain semantics, the vase is not destroyed. Perhaps in this world the vase no longer exists. But the vase is not dependent on this world. The vase actually resides within a wider domain (because the domain of objects is fixed in all possible worlds).[5] The vase exists in another possible world covered by the domain. Likewise, Kirk does not affect the domain of objects simply because he does not exist in the actual world. He just is in some other possible world.

Section II: Objections to Priest's view: The Quine & Lewis Arguments

Quine's paper, *On What There Is*, is considered one of the most influential attacks against Meinongianism. Although Priest's view varies from Meinongianism, in *Towards Non-Being* Priest responds to some of Quine's arguments. I will give one of Quine's arguments, and then I will provide Priest's response. I will also present one of Lewis's arguments against Noneism as well.

Although Lewis's argument is directed at Sylvan, I am going to direct Lewis's argument towards Priest's Noneism as well. I will also give Priest's response to Lewis.

Quine believes that all that exists is what can be quantified over. As mentioned previously, he argues that Meinong obfuscates the word 'exist.' He writes:

> Despite his espousal of unactualized possibles, he limits the word 'existence' to actuality….We have all been prone to say, in our common-sense usage of 'exist,' that Pegasus does not exist, meaning simply that there is no such entity at all. If Pegasus existed he would indeed be in space and time, but only because the word 'Pegasus' has spatio-temporal connotations, and not because 'exists' has spatio-temporal connotations. If spatio-temporal reference is lacking when we affirm the existence of the cube root 27, this is simply because a cube root is not a spatio-temporal kind of thing, and because we are being ambiguous in our use of 'exist.' (Quine 1953, p. 3)

Quine is attacking the Noneist view of predication, namely that 'existence' is considered a predicate on this view. According to the Noneist, Pegasus *is* in a possible world that realizes the situation. And in that possible world, Pegasus contains the property of existing, among other such properties as 'can fly' and 'has hooves.' But Quine remarks that this is not what most people mean by nonexistence; we do not mean that Pegasus *is* in a possible world. Thus Quine writes, "The only way I know of coping with this obfuscation of issues is to *give* [Noneism] the word 'exist'. I'll try not to use it again; I still have 'is' (Quine 1953, p. 3).

Priest responds that Quine is not targeting Noneism. He writes, "Many of the Meinongian objects <u>are</u> not" (Priest 2005, p. 108). He asserts that Quine's argument falls short. For the Noneist, to be and to exist are identical. This is just to say that Priest's view is different from Meinong's view.

Priest then raises Quine's argument regarding quantification. He writes: "It is in this part of the paper that Quine proposes

his famous thesis about existence and quantification (p. 13): 'To be assumed as an entity is, purely and simply, to be reckoned as the value of a variable.' Or, as it is often more pithily put: to be is to be the value of a bound variable" (Priest 2005, p. 110). In other words, Quine critiques the Noneist view of quantification, in which constant-domain semantics is endorsed. Priest asserts that no argument is given for this view. However, Lewis gives an argument here. Lewis writes, "there is only one kind of quantification" (Lewis 1990, p. 24). The Noneist, however, separates quantification into two kinds: *existential quantification* (or existentially loaded quantification) and *existentially neutral* (or 'particular') quantification.[6] Lewis asserts that 'existence' is reserved for existentially loaded quantification; this is restricted to things that exist unparadoxically—i.e. concrete objects. For the existentially neutral quantification, "there are" and "some" are to be used neutrally; thus, we can quantify over all things whether or not the objects exist (Lewis 1990, p. 25). The Noneist thus professes that 'some things do not exist' (Lewis 1990, p. 25).

According to Lewis, the several idioms of existential quantification proposed mean exactly the same thing. Lewis writes, "It does not matter whether you say 'Some things are donkeys'or 'There are donkeys' or 'Donkeys exist'—you mean exactly the same thing whichever way you say it" (Lewis 1990, pp. 24-25). The Noneist separates "∃x" from the standard view: in his own paper, Priest uses a new notation so as to indicate a different reading of the particular, or neutral, quantifier. Rather than read the existential quantifier as, "there exists an *x* such that," Priest asserts it should be read as "some *x* is such that."

Lewis faults the Noneist for quantifying without quantifying. He writes: "For when he quantifies neutrally he is not quantifying in the one and only way there is to quantify, since ex hypothesi the one way is the loaded way" (Lewis 1990, p. 27). Lewis asserts that dividing quantification into either a loaded or neutral category is unintelligible; he cites Lycan, who deems Meinongian quantification "literally gibberish or mere noise" (Lewis 1990, p. 27). He asserts that the Noneist's 'existentially neutral' quantification is

really the same as Lewis's (and others who aren't Noneists) one kind of quantification.

Moreover, Lewis criticizes the Noneist for quantifying over everything that exists—just as the he does. Lewis writes: "He affirms the existence of all the controversial entities (as we may call them). He does not join us when we dodge questions about some of these alleged entities by denying that they exist" (Lewis 1990, p. 29). Thus according to Lewis, Sylvan (and also Priest) is really an Allist. An Allist simply asserts that *all* the controversial entities exist. This is in contrast to the Noneist doctrine that holds *none* of the controversial entities exist. But, according to Lewis, Priest's existential commitments commit him to Allism. Lewis argues that when the Noneist restricts his quantifier to trees, it is easily understandable because Lewis (and other Non-Noneists) also distinguishes between trees and non-trees (Lewis 1990, p. 30). But when the Noneist loads his quantifiers, "he restricts them to the entities which, he says, 'exist'—and this is not understandable" (Lewis 1990, p. 30). Lewis makes no such distinction among entities. Lewis writes, "if 'existence' is understood so that it can be a substantive thesis that only some of the things there are exist—or, for that matter, so that it can be a substantive thesis that everything exists—we will have none of it" (Lewis 1990, p. 31). Therefore, according to Lewis, Noneism should be rejected.

Priest gives his response to this critique in a paper titled, "Against Against NonBeing." At first he responds by arguing that Lewis's critique is methodologically flawed. He writes, "There is absolutely no reason why, in a dispute between noneists and Quineans, everything said by one side must be translated into terms intelligible to the other" (Priest 2008, p. 251). He argues that it is not expected that notions of Special Theory of Relativity need to be translated in Newtonian Physics, or that Marxian economics are translatable into Keynesian economic terms. Therefore, to expect a Quinean to use the same language as a Noneist is unreasonable. Moreover, he writes: "Most Quineans I know seem to me to understand the noneist language game perfectly outside the seminar room" (Priest 2008, p. 251).

He concludes by arguing that the Noneist has a richer conceptual scheme than the Quinenan. Both sides intend "$\exists x$" to mean "there exists"; yet, as Quineans cannot express the Noneist particular quantifier, their schema is not as rich. He writes:

> Richer conceptual schemas are methodologically desirable. Hence Translation 1 is to be preferred. This does not commit a Quinean to being a noneist. They may hold $\forall x \exists x$ (i.e. $\forall x \exists y\ x = y$). In this language they can actually *express* their view, and not just presuppose it. Noneists do not, therefore have a bloated ontology—that is, an overgenerous view of what is/exists. (Priest 2011, p. 252)

Hence when Lewis writes, "he affirms the existence of all the controversial," it is precisely that which Priest denies. He is quantifying over nonexistent things. And this does not seem challenging to Priest. Nonexistent objects are simply realized in possible or impossible worlds.[7] In *Towards Non-Being*, Priest argues that Quineans have a bloated universe, for the Quinean asserts *all* the controversial objects exist. Neutral quantification is useful when objects in the domain do not exist—when objects are not concrete. The Noneist's universe is more exacting; they affirm only what exists. Nonexistent objects *are* not in the actual world; but there are worlds that realize fictional objects.

Section III: How We Should Understand Nonexistent Objects

Quine and Lewis both object to Priest's view, and Priest responds. But despite Priest's response that "outside the seminar room" his view is perfectly understandable, it really is not. I am arguing against the thesis that a fictional object like Kirk can have properties *and* not exist. The supposition that nonexistent entities can have properties fails to capture the true difference between fiction and reality. Existence is required for an object to have properties. The Quine/Lewis view argues flat incomprehension—it is just incomprehensible for an object to have properties and not exist. I

too am unconvinced by the application of an existential predicate.

Priest's response to Lewis is inadequate. He does not address the metaphysics of his view when arguing that the Noneist language is perfectly understandable "outside the seminar room." The metaphysics of his view seems so similar to that of a Platonist, in that fictional objects do not exist (they are not spatiotemporal things), yet these objects have properties. It is rather bizarre to believe both that Pegasus does not exist and in a world that realizes the situation Pegasus has the property of existing.[8]

The right way to think about fictional objects is that they do exist in the actual world, but as abstract objects (non-mental, non-physical things). *Prima facie*, existence is required for an object to have properties. This is what Lewis terms an Allist view: *all* the controversial objects exist. Moreover, truth conditions for sentences containing fictional names are determined *according to the story*. We need not rely on possible worlds as Priest does. I endorse a version of Kripke's view as it relates to fictional objects.[9] On this view, the name of a fictional object is a rigid non-designator—it refers to nothing at all. Thus, a sentence containing a fictional name, like the name 'Kirk,' refers to nothing with respect to every possible world (refers to nothing in any possible world). On one usage, it has no referent, but on this usage there is a *pretend* use of the name 'Kirk'. 'Kirk' only pretends to name a human being, but it in fact names nothing at all. The second use, the non-pretend use, arises when someone is speaking about the work of fiction from outside the work. In its non-pretend use, the name 'James T. Kirk' genuinely refers to a particular abstract object. We can call the first use of 'Kirk' the metaphysical use, and we can call the second use a semantic use. And we can disambiguate 'Kirk' by speaking of 'Kirk_1' (metaphysical) and 'Kirk_2' (semantic). And yet neither names a real object. On this view, 'Kirk_1' has a thoroughly nonreferring use; 'Kirk_2' names an abstract object.

Furthermore, we can distinguish between two kinds of sentences on Kripke's view: *fictional* sentences and *meta-fictional* sentences. Meta-fictional sentences contain the operator (according

to the story), either explicitly or implicitly. Fictional sentences are used to make a claim from within the fiction. Fictional sentences can be considered true or false in an extended sense—truth *in the fiction*. Thus, a sentence is evaluated for its truth conditions as it pertains to the fiction. The statement, "Kirk is captain of the enterprise" is true according to the fiction. However, should the name be used in the manner of Kirk_2 (the semantic use), then the aforementioned sentence is false because then the statement makes a claim about reality rather than a claim about the fiction. Lastly, predicates (i.e. a verb-phrase of the kind 'captains a Star Ship') also have an extended sense on which they can be evaluated for their truth conditions. The predicate 'captains a Star Ship' applies to the abstract entity in question only when the corresponding fictional entity captains a Star Ship.

However, I think that fictional, abstract objects have an additional property: they are mind-dependent. An object is mind-dependent if and only if it is an object of some mental state or process, and that object could not exist independently of minds. Hence, the properties of an object, Φ, are dependent upon an existing entity, ε, only if it cannot be the case that Φ's properties differ without intervention from entity ε. In other words, Φ's properties cannot change without ε changing them.

For example, the fictional abstract object James T. Kirk cannot change its properties without an existing entity (perhaps a writer on the show *Star Trek*) causing those changes. Fictional objects are dependent upon existent objects—if minds do not exist, then fictional objects do not exist. As such, fictional abstract objects causally depend on existent entities. For object x to causally depend on y, y must interact with x in such a way that there is a direct correlation between y and x-properties. Roddenberry, as he is formulating the character Kirk, amalgamates Kirk-properties, such as "is a captain" and "is even-tempered." In this sense, he *creates* Kirk. But *creates* here does not mean *brings into existence*; instead, *creates* means something like *formulates*, in which Roddenberry formulates Kirk by pulling together Kirk-properties. Thus, Kirk does not spring into existence through Roddenber-

ry's imagination. Rather, Kirk-properties exist, and now, due to Roddenberry's fiddling, we have the character known today as Kirk.

Fictional objects are therefore mind-dependent, for they are characterized as having certain properties *by* existing entities. Actually existing conscious entities (namely, people—unless animals have a richer consciousness than currently surmised) give rise to the possible worlds in which fictional objects are realized through their intentions, utterances, and imaginations. Although Kirk exists outside of Roddenberry's mind, he is created by that mind, and his properties are thus dependent on how Roddenberry chooses to characterize Kirk.

The idea that fictional abstract objects are mind-dependent is rather controversial in discussions concerning abstract objects. One foreseeable objection may be the following: from this definition of mind-dependence, it seems that we can classify not only abstract objects as mind-dependent, but also concrete objects as well. Hence buildings are mind-dependent, as buildings could not exist without minds to create them.

But within the parameters of my view of mind-dependence, an object is mind-dependent if and only if it could not exist independently of minds. And surely we would classify a currently existing concrete object, like a building, as having the property of existence even if no minds are there to reinforce it. In other words, if all minds were to suddenly disappear from the actual world, currently existing buildings would persist. But fictional objects would not. Existing fictional, abstract objects would not persist if all the minds disappeared. If the active world of minds ceased to exist, would Roddenberry's formulation of Kirk persist? It seems clear that the answer is no. Kirk-properties may persist, but Roddenberry-formulated-Kirk would not persist. Thus, these kinds of abstract objects are mind-dependent.

Conclusion

I have argued that fictional objects are abstract objects and that

they are mind-dependent. I remain unconvinced by Priest's view that existence is a predicate. I share the principal concern of Quine and Lewis that existential predication is incomprehensible. I have argued that Priest's view of existential predication is incomprehensible.

I am aware that the matter of what fictional objects are remains open to debate. Certainly my argument here does not lay any outstanding concerns to rest. But perhaps this paper provides one explanation of what we say when we utter sentences like "James T. Kirk is captain of the Enterprise" or "Sherlock Holmes is a detective." At least we have an argument that supports the idea that fictional abstract objects are not nonexistent; rather, fictional abstract objects do exist and they are mind-dependent.

Notes

1. The term 'actual' designates this world, the world in which you and I currently exist.
2. There is no clear definition of an abstract object because the question of what it is to be an abstract object is widely debated in contemporary philosophical discourse. However, according to the Stanford Encyclopedia of Philosophy, "it is universally acknowledged that numbers and the other objects of pure mathematics are abstract (if they exist), whereas rocks and trees and human beings are concrete. Some clear cases of abstracta are classes, propositions, concepts, the letter 'A', and Dante's *Inferno*. Some clear cases of concreta are stars, protons, electromagnetic fields, the chalk tokens of the letter 'A' written on a certain blackboard, and James Joyce's copy of Dante's *Inferno*." For more information see Rosen, Gideon, "Abstract Objects", *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2014/entries/abstract-objects/>.
3. Intensional states assign sets of objects to predicates. For example, the semantic value (or meaning) of 'green' maps each possible world to the set of things in that world that are green. See the following entry in *The Stanford Encyclopida of Philosophy* for a discussion on properties and intensional states: Swoyer, Chris and Orilia, Francesco, "Properties", *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2014/entries/properties/>
4. For the purpose of our discussion, I am not relying Priest's more technical explanations of how these worlds are manifested.
5. I believe this is largely Sylvan's Existentially neutral 'particular' quantification. He asserts that "there are" and "some" are to be used neutrally; thus

quantification over all things that exist. Sylvan, Richard (1980) *Exploring Meinong's Jungle and Beyond: An Investigation of Noneism and the Theory of Items*, (Australian National University) pp. 174-80 and elsewhere), as cited in Lewis 1990, pp. 25

6. Sylvan, Richard. (1980) *Exploring Meinong's Jungle and Beyond: An Investigation of Noneism and the Theory of Items* (Australian National University), pp. 174-80 and elsewhere, as cited in Lewis 1990, pp. 25
7. Priest is a realist about possible worlds.
8. This response to Priest came from a discussion with Mark Balaguer. I want to thank him for helping me respond to Priest in this way.
9. My understanding of Kripke's view comes from Salmon's article published in 1998.

Bibliography

Lewis, David. (1990) "Noneism or allism?," *Mind* 99, pp. 23-31

Priest, Graham. (2011) "Against against nonbeing," *The Review of Symbolic Logic* 2, pp. 237-253

_____. (2011) "Creating non-existents," *Truth in Fiction*, pp. 107-118

_____. (2008) "Précis of towards non-being" *Philosophy and Phenomenological Research* 1, pp. 185-190

_____. (2005) *Towards Non-Being: The Logic and Metaphysics of Intentionality* (Oxford: Clarendon Press)

Quine, W.V.O. (1953) *From a Logical Point of View* (Cambridge, MA: Harvard University Press)

Salmon, N. (1998). "Nonexistence," *Noûs* 32(3), pp. 277-319.

Meta-Ontological Disputes

Jonathan Fry

Introduction

Ontology can be described as the study of what exists. There are many ontological disputes, such as the dispute over whether or not composite objects (such as tables) exist. Another separate and more general debate concerns whether or not ontological questions have determinate answers and whether or not these answers are trivial. This is referred to as meta-ontology. Many philosophers have contributed to the meta-ontological debate, providing a wide array of theories. David Chalmers has categorized some of the positions held among meta-ontologists, drawing a distinction between ontological realists and anti-realists. Ontological realists claim there are objective answers to questions of existence, while anti-realists claim there are no such answers. Those who say that there may be objective answers to ontological questions, but claim that they are simply uninteresting and not worth fighting over, are known as deflationists. Deflationists fall under the ontological realist category. Chalmers describes them as *lightweight* ontological realists, while those who believe the answers to ontological questions are actually substantive are *heavyweight* ontological realists.

I will mainly address two different meta-ontologists' views: Karen Bennett's and Amie Thomasson's. Bennett is in fact an ontological realist, but she provides a view that is neutral with respect to the ontological realist/anti-realist divide. Her view is that we are unable to determine which side of a particular ontological debate is more justified because they both encounter the same problems. Thomasson would be classified as a deflationist or lightweight ontological realist, but that simple classification does not adequately represent her view, as will be shown. She believes that,

at least in the particular debate about composition to be discussed, the two opposing sides are merely talking past each other: that there really is not a debate to be had. In the rest of this paper, I will summarize Bennett's epistemicism and Thomasson's semanticism as they present them, providing an analysis of each view in turn. I will argue that Bennett's position is untenable because, contrary to her claim, one position within the debate about composition—i.e., nihilism—is more justified than the other position, universalism. I will argue that Thomasson's position is also untenable because she does not allow for both sides of the debate. Finally, concluding that the only positions left are anti-realism and heavyweight realism, I will briefly consider and respond to arguments for both, concluding that realism is the better choice.

1. Bennett's Epistemicism

1.1 Explanation

Karen Bennett, in her paper, "Composition, Colocation, and Metaontology," characterizes three different versions of dismissivism, her name for the general view that there is something wrong with many metaphysical debates, specifically ontological debates that have to do with questions of the existence of a type of thing. The first version of dismissivism she brings up is ontological anti-realism, which will be discussed later on in the paper. The second is *semanticism*, of which Thomasson's view is a version. The third is Bennett's own brand of dismissivism, *epistemicism*. She does not discuss anti-realism much; instead, Bennett argues against semanticism, the view that these metaphysical disputes are merely verbal ones, and then argues for epistemicism, claiming that there is little justification for believing one side over the other within ontological debates.

One of the most important things Bennett explains in her paper is her methodology. She decides that it is not correct to lump all metaphysical or even just ontological debates into one descriptive category. It may be the case that certain debates are of a different nature than others, and that the conclusions one comes

to about one type of debate might not apply to others. For this reason, Bennett focuses on specific ontological debates, problems concerning constitution and composition. A metaphysical puzzle concerning constitution is the *statue/lump problem*, which can be described as a problem of multiple seemingly distinct objects being located in the same region of space. For instance, imagine a statue called 'Nosmashy'. Now, take the lump of clay that Nosmashy is made up of and call it 'Smashy'. The real problem can be seen when analyzing what would happen were one to smash Nosmashy/Smashy. It seems that the statue would cease to exist, while the lump would remain; so Nosmashy and Smashy have at least one different property, namely *smashing-survivability*. Therefore, by Leibniz's Law[1], Nosmashy and Smashy are distinct objects; but (before the smashing) they reside in the same location at the same time; hence the problem.

The problem of composition is an ontological debate over just how many objects actually exist. When two or more objects make up a third distinct object that is the combination of the first two, we say the first two objects compose the third or, equivalently, that the third object is the composite of the first two objects. The question is: *when does composition occur?* There are an infinite number of ways to answer such a question, but a good answer should be principled. There are two extreme, principled answers to the composition question: *never* and *always*. The former answer corresponds to what I would call reasoning from the negative side; one could assert that there definitely does not exist the object composed of the planet Mars and my computer and then reason from this that composition does not occur at all. This may lead to unwanted results, however, like the denial of the existence of tables. Another way to approach the problem is from the positive side, in which you assert that tables definitely exist, then reason that if composition occurs in that case, then it must occur all the time, leading to the denial of the first assertion of the negative side, or the assertion of the claim that there is an object composed of Mars and my computer. The first of these views, the view that composition never occurs, can be called (mereolog-

ical) nihilism, while the view that composition always occurs can be called (mereological) universalism.[2] Bennett calls the person who believes there is just one object in the statue/lump example a *one-thinger* and the person who thinks there is more than just one object a *multi-thinger*. She calls the nihilist a *nihilist* and the universalist a *believer*. Finally, she labels the one-thinger and the nihilist low-ontologists and the multi-thinger and the universalist high-ontologists (for obvious reasons). Those final distinctions come into play in her reasoning for her dismissivist position.

Bennett reasons that both cases, constitution and composition, can be collapsed into a discussion of composition alone. Rather than talking about how many objects are constituted in the statue/lump case, we can simply ask how many objects the particles of clay compose. The one-thinger would claim that one object is composed; the multi-thinger, at least two; the nihilist, none; and the believer, at least one. So, it seems the nihilist avoids the statue/lump problem altogether by not believing in statues or lumps. The believer, on the other hand, must first decide whether she is a one-thinger or a multi-thinger. Then, if she is a one-thinger, she must explain why there seems to be at least two distinct objects in the same space and time, based upon the argument from Leibniz's Law; and, if she is a multi-thinger, she must explain how there can be multiple objects in the same location. It is interesting to note that the same ontological perplexities at play in the statue/lump co-location problem arise in the general problem of the existence of composites since the nihilist must explain why there seem to be composite objects, like tables, while the believer (whether one-thinger or multi-thinger) must explain the exact same problem of *how there can be* multiple objects in the same space and time yet again. This is because the believer believes in both simples (the smallest parts of things that do not have parts themselves) arranged table-wise *and* tables. These multi-level perplexities, I think, provide prima facie reason for the nihilist position (supposing meta-ontological realism), but more on this later. As shown, it is useful to narrow down the cases and simply talk about composition (which I will do for the rest of the paper as I follow

Bennett's lead in the assumption that, in order to do metametaphysics, we must first do metaphysics). I will continue with the case of composition as the standard ontological debate of meta-ontological discussion and reapply Bennett's low-ontologist/high-ontologist distinction to the question of whether or not tables exist. The low-ontologist in this case would be the denier of table existence and the high-ontologist would accept table existence.

Bennett points out that the high-ontologist will generally boast in the expressive power with low ideological cost of her view, while downplaying her excess ontology. Contrarily, the low-ontologist will generally boast in the simplicity of her few-object conception of the universe, while up-playing her expressive power, repositioning the cost of a high ontology to a high ideology. Bennett describes these two different tactics as a common phenomenon of difference-minimizing. Each side wants to make it seem like their view is not so outlandish. She moves from this observation that both sides seem to have the same amount of problems, simply shifting the blame of the difficulties to different areas—a situation she analogizes as a bump under a carpet being pushed down, just to reappear elsewhere—to her epistemic stance. She reasons that there just are no sufficient grounds for choosing one side over the other; so it is pointless to continue to fight over ontological questions (at least composition). Bennett is sure to differentiate her view from typical deflationary views: "My claim, then, is not that work on the metaphysics of material objects is pointless, but rather that *we have more or less done it already*" (Bennett 2009, p. 73).

1.2 Evaluation

Although Bennett is a realist when it comes to ontological debates, and epistemicism, as she defines it, includes a component of realism, she actually argues for a position that is neutral on the realism/anti-realism dichotomy. It is compatible with either. One could hold that there are objective answers to existence disputes or that there are not, but either way think that there is not enough evidence to be justified in beliefs about the existence of tables, for example.

The problem with accepting epistemicism is evident if the big picture is considered. One is presumably supposed to give up one's low-ontologist or high ontologist beliefs and throw in the towel simply because it is pointed out that his opponent has good reason to believe what she believes! Usually a person will take and remain on one side in a dispute because they feel they are more justified on that side than on the other. Bennett's argument is for the position that there is no point in fighting for one ontological view over another *anymore* because the work has already been done and neither side is more justified than the other. The problem is just that, for people who defend their ontological theory, the point in arguing simply is that they think they are more justified than their opponents. For instance, I have already stated that the perplexities of the statue/lump and composition problems, taken together, provide a prima facie reason for nihilism. The nihilist seems to avoid the continually rising problem of there being too many objects in the same space at the same time. This problem should not be minimized. The believer claims that there is some *extra* entity besides the simples that compose it (no matter how many composite objects she says there are). I am not sure what this extra entity *is* and what it *does* over and above what the simples are and do; these extra objects are *spooky*.[3]

One might not think that I am being fair to Bennett; her view is *epistemic* in that she does not think we have access to the relevant information necessary to determine which side of the composition debate, the low-ontologist or the high-ontologist, is correct.[4] If that is the case, then my arguing for the low-ontologist side gets me nowhere. However, Bennett makes it clear that her epistemic dismissivism relies heavily on two claims: that the low-ontologist's position is not clearly simpler than the high-ontologist's and that both views suffer from the same problems.[5] She bases her argument for the first claim on the notion that, just as the high-ontologist needs to downplay her excess ontology, the low-ontologist needs to downplay his excess ideology. The table example might make this exchange of downplaying clearer. The high-ontologist has her table to work with; she can usefully talk

about it as if it is just another object because she believes it is, but, as pointed out previously, it turns out that the notion of a table as an object distinct from the parts that make it up is strange and must be explained. The low-ontologist escapes having to explain such strange objects in his view, but he must explain how and why we talk about such objects as if they were real. This involves an explanation of sentence truth-values, meaning, and reference, most importantly, for composite objects that show up in the statements of important scientific theories. The problem with all this discussion of down-playing ontology as opposed to ideology is just that the low-ontologist's conception does seem simpler. Truth, meaning, and reference (all concepts I consider to be more related to ideology than ontology) seem to be the types of things that can be tinkered with and explained; whereas ontology, whether or not something is, does not seem to be the type of thing that can be down-played. Extra-physical objects cannot just be explained away. This is why nihilism (low-ontology) is clearly simpler than universalism (high-ontology). Of course, the notion of simplicity is (excuse the pun) complex and has been a topic of philosophical discussion for a long time. This paper is already too full to go into more detail.

A further reason for not throwing in the towel in the ontological debate, even if opposing positions in the debate are each not giving up any ground, could be that one thinks their view has important implications for other areas of inquiry. I do not think that someone would be getting lost in a pointless endeavor if they were to pursue their ends unflinchingly in an ontological debate. It may seem pointless; after all, we are still going to continue to trust that our dinnerware will be supported by our dining tables should the nihilist be shown to be correct. But there is reason to believe that people's metaphysical views (even if metaphysics turns out to all be bunk and even if they are unaware that they even have a metaphysical view) can shape the way they look at the universe. This can affect an individual in, say, their scientific pursuits of knowledge; and there is no denying the powerful influence of science on humanity and the world. This connection between a

metaphysical view and practical effects in the lives of people may seem like a stretch, but science is driven by scientific theories, and scientific theories are laden with metaphysical assumptions.[6] Much more could be said about this.

However, there is something to Bennett's bump-under-a-rug analogy, especially considering her arguments that the same problems that plague the high-ontologist can be reformulated for the low-ontologist.[7] Epistemicism could alternatively be looked at as a call to ontologists: rather than debating ontological questions endlessly, they should try to develop a method for determining which ontological theory is superior to others in a given debate (maybe by defining a notion of simplicity). Bennett admits that her view would be dismantled if it could be shown that the low-ontologist side *is* more justified because admitting fewer objects into a conception of the universe is a higher priority indication of a simpler, better theory.[8] It could also turn out that the high-ontologists have better justification for their position because a simpler ideology is what is necessary for a better theory.

One may still think that there is something else fundamentally wrong with ontological debates separate from Bennett's epistemic worry, including the worry that the debaters are simply talking past one another. This is Thomasson's position, which will be discussed next. If I can show that Thomasson's semanticism (and therefore other similar deflationary positions) is untenable, then the only other serious options are heavyweight ontological realism and ontological anti-realism.

2. Thomasson's Semanticism

2.1 Explanation

Amie Thomasson, in her paper, "Answerable and Unanswerable Questions," develops and argues for a view about reference that provides a basis for answering some metaphysical questions. She also asserts that her view of reference provides a basis for determining when some metaphysical questions are unanswerable, arguing that many important metaphysical debates, specifically

ones concerning ontology, are meaningless or pointless.

Thomasson believes generally that a causal theory of reference is correct, but reasons that if reference were determined purely causally, then reference failure would occur more frequently than it does. There must be some reason why things are not more often difficult to ontologically disambiguate, and since that reason does not come from *outside* of us as part of the purely causal part of reference, it must come from *inside of us by conceptual analysis*. This term-disambiguating conceptual analysis comes in two forms: identifying 'frame-level application conditions' and identifying 'frame-level coapplication conditions'. Identifying frame-level application conditions involves learning and using a term, and determining when it applies and when it does not. Thomasson uses the example of learning the difference of when the term 'lump' applies and when the term 'statue' applies. We analyze our own concepts to determine that 'lump' applies just when there is a cohesive mass of stuff, but 'statue' applies when a figure was intentionally sculpted. Identifying coapplication conditions involves determining when a term can apply to an object again after it's been applied before. Coapplication conditions, then, have to do with persistence conditions. Thomasson defines sortals as general nouns that involve these application and coapplication conditions. Overall, these conditions disambiguate which ontological category an entity belongs to (whether or not something is, say, an animal, or an inanimate object). So, she sets up her theory of reference as a hybrid theory, with general ontological categories of terms we use being disambiguated through conceptual analysis, but specific referents being determined causally.

Her theory leads her to make the claim that we can determine whether or not something exists by deciding whether or not a term refers to it. This is done via a two-step process. First, we decide what the frame-level application conditions are for a certain term to refer. Then we decide whether or not those application conditions are met. Thomasson applies this process to the mereological question of whether or not tables exist. She argues that both the nihilist and the believer would have similar frame-

level application conditions for when tables exist and when particles are arranged table-wise, so that it seems the nihilist really should admit that tables exist.

Thomasson anticipates the nihilist's response that even if certain application conditions were met, not all of them would be. There would still be the condition of being an object distinct from the particles arranged table-wise, a condition which the nihilist would deny is met. Thomasson argues that this response pushes the debate to the general level of existence questions. She characterizes general existence questions as disagreements about the uses of the general terms 'object' and 'thing'. She divides uses of these terms into three categories: sortal uses, covering uses, and the neutral use. She then argues that the sortal and covering uses of the terms still do not ground the deep metaphysical nature of the disputes. They are still merely verbal because the debate is solely about either the application conditions being met or what the application conditions really are. The only chance to revive the substantial nature of the ontological debates is for there to be a neutral use of 'object' or 'thing'. Thomasson claims this is not going to work. A neutral sense of 'object' or 'thing' would imply that there are no application conditions because, if there were, then the debates would still be merely verbal as they were with the sortal and covering uses. Without these application conditions, though, there would be no way to evaluate the truth of whether or not the terms apply; so the question of whether or not something existed would be unanswerable. So, Thomasson's view is that either existence debates (at least ones having to do with composition and the like) are shallow, merely verbal disputes—not substantial ones—or they are debates over questions which are unanswerable.

2.2 Evaluation

There is a fundamental problem with Thomasson's semanticism. As presented, there is absolutely no way that the real ontological debates worth talking about involve the use of 'object' and 'thing' as sortals or coverers. It is just not true that the believer, who

believes there are tables, and the nihilist, who believes there are no tables, are merely talking past each other. To assert that they are talking past each other is to deny their positions from the start. Asserting that all that it takes for there to be a table is for there to be particles arranged table-wise does not even allow the believer's real position into the debate. The believer does not just think that there are tables because there are particles arranged table-wise; she believes that the particles arranged table-wise compose a whole separate object, a table. The nihilist's real position is also being overlooked. She admits that there are particles arranged table-wise, but denies the existence of tables. The debate cannot be as Thomasson says it is because the relationship between simples and a composite object cannot be identity. Identity is a one-to-one relationship, not a many-to-one. The true believer accepts this. Even if certain participants in the debate mean what Thomasson says they should mean, namely, that all it takes for there to be tables, is for there to be simples arranged table-wise, we could still resurrect the deep metaphysical debate on behalf of the true nihilist and true believer. Thomasson is correct when she allows that the debate is about tables as objects in the neutral sense. But she says, in the neutral sense, the question of whether or not a table is a thing or an object is unanswerable.

This all leads me to believe that either Thomasson is a closet nihilist (at least about tables), which would make her a meta-ontological realist, or she is an epistemicist like Bennett, or she is a true anti-realist. Of course, she would deny any of these labels, but if she really believes that all it takes for there to be a table is for there to be simples arranged table-wise, then she does not believe in the extra, spooky entities the believer posits; and, therefore, she is actually a nihilist about tables. However, if we allow that the debate involves the neutral sense of 'object' (which it must, if we are talking about the deep metaphysical debate), and she holds true to her claim that the questions become unanswerable, then either she shares Bennett's view that the questions are unanswerable because the answers are not epistemically available to us or she is an ontological anti-realist and thinks that the questions do

not even make sense.[9]

With Bennett's epistemic dismissivism and the verbal-deflationary views of Thomasson's sort off the table, we are approaching the decision between heavyweight realism and full-fledged anti-realism. I consider my main objective in this paper accomplished: narrowing the field down to just the contradictory positions of realism, the view that there are objective answers to the question of composition, and anti-realism, the view that there are not. The last section of this paper will be devoted to an exposition of the debate directly between these two opposing views, each of which I think it is difficult to argue for, as I hope will be shown.

3. Ontological Anti-realism and Realism

It is difficult to know what to say to an ontological anti-realist. Bennett expresses similar difficulties:

> I do not know how exactly to argue against it [anti-realism], and I am not entirely sure what it means. 'There are *F*s' might be vague or ambiguous in some way, in which case the *unprecisified* sentence might not have a determinate truth-value. But I am not entirely sure how it could be that a *precisified* version of the sentence does not have a truth-value. (Bennett 2009, p. 40)

This quote could be read as an argument against anti-realism, but I think we should take Bennett at her word that she does not know how to argue against it, and she is not entirely sure what it means. It is difficult to make sense of anti-realism, and I think that is because of the depth of the position. Ted Sider describes a view that I take to be anti-realism as "reject[ing] the notion of objective structure altogether" (Sider 2009, p. 418). He regards holding this position as "unthinkable."

Chalmers provides a sketch of an argument, the *conceivability argument*, against ontological realism (which I take to be an argument for anti-realism) that mentions a "nondefective abso-

lute quantifier" that seems to be akin to Thomasson's neutral sense of 'object':

> If there is a nondefective absolute quantifier, then both nihilism and universalism are conceivable: neither of them can be reduced to contradiction by a priori reasoning. But what cannot be ruled out a priori is possible… Nihilism and universalism are not both possible. So there is no nondefective absolute quantifier. (Chalmers 2009, p. 104)

Chalmers says that more needs to be said for this argument to be taken seriously, but a quick argument deserves at least a quick response. The conceivability argument makes use of one logical notion of possibility to refute the application of another logical notion of quantification to the structure of the world. Metaphysical possibilities, if they are different from logical possibilities, are difficult to determine. In fact, the determination of metaphysical possibilities requires the use of intuition, and it is not clear that nihilism and universalism are not both possible.

An argument for ontological anti-realism in the same vein of Chalmer's conceivability argument (because it makes use of the notion of possibility and necessity, but is handled in much more depth) is the topic of Mark Balaguer's paper, "Why The Debate About Composition is Factually Empty (Or Why There's No Fact of the Matter Whether Anything Exists)." The argument is complicated, but I think a fair synopsis of the strategy of the argument is to reveal that ontological realism is ultimately an unfounded assumption, and then to place the burden of proof on the realist. I think Balaguer does justice to the debate by keeping us wary about our metaphysical assumptions. Sider expresses this wariness:

> I think we should remember something that often gets lost in these [meta-ontological] debates. *Everyone* faces the question of what is 'real' and what is the mere projection of our conceptual apparatus, of which issues are substantive and which are 'mere bookkeeping'. This is true within

> science as well as philosophy; one must decide when competing scientific theories are mere notational variants… Unless one is prepared to take the verificationist's easy way out, and say that 'theories are the same when empirically equivalent', one must face difficult questions about where to draw the line between objective structure and conceptual projection. (Sider 2009, pp. 416-417)

I take all of these considerations to be good reasons for not dismissing anti-realism off-hand. It seems that some types of things exist, while others do not, and it also seems that there should be some way of determining which types exist and which do not. However, maybe it is the case that when we say things like "composite tables do not exist, but the simples that supposedly compose them do," we are already mistakenly assuming a certain underlying structure of the world. The question is, "in the sense that we have been talking about, does the universe embody the structure implied by a neutral use of the word 'object' or a 'nondefective absolute quantifier' apart from our conceptual apparatus?" While I appreciate the difficulty and complexity of the issue, I agree with Sider that we eventually must draw the metaphorical line between objective structure and conceptual projection. Sider bases his decision of where to draw the line on the methodology of W.V.O Quine:

> [T]he ontological realist can give a pretty convincing argument for his choice of where to draw the line. Quine's (1948) criterion for ontological commitment is good as far as it goes: believe in those entities that your best theory says exists. But in trying to decide how much structure there is in the world, I can think of no better strategy than this extension of Quine's criterion: believe in as much structure as your best theory of the world posits. (Sider 2009, p. 417)

Our best theories of the world are those proposed by science. Improvements in science have led to huge advancements in tech-

nology, which have (arguably) improved our quality of life and our understanding of the world. There is a trend within science to explain phenomena by recourse to the existence of smaller and smaller entities. The properties of these entities are used, in many cases, to explain the properties of larger entities, such that nothing more is required to describe phenomena in general than what is already explained by the smaller entities. The extent to which this process can and has been carried out is a controversial topic within science and the philosophy of science. However, there is no doubt that the trend is there and that in some cases, recourse to larger entities can be replaced with recourse to smaller ones. Since this is the case, and because I am optimistic that it will continue to be the case more and more as our best scientific theories progress, I agree with Sider that we should believe in as much structure as our best theories posit. Our scientific theories posit that there is something rather than nothing and that bigger things can generally be explained in terms of smaller things. This leads me to accept realism when it comes to ontological debates and to attribute structure to that reality based upon a nihilistic approach. All I mean by this is that we should only accept the existence of entities that are not explainable in terms of any smaller parts. This provides as clear of a place to draw the line between objective structure and conceptual projection as I can think of, and, as Sider suggests, the line must be drawn.

Conclusion

In this paper, I have explained four different positions in meta-ontological debates insofar as they relate to the debate over the existence of composite objects. I have argued against Karen Bennett's view that answers are epistemically unavailable to us and Thomasson's view that the debates are merely verbal. This led into a discussion of the debate between the only remaining meta-ontological views, heavyweight realism and anti-realism, in which I briefly discussed the opposing views and offered an argument for heavyweight realism on basis of the success of science.

Notes

1. Leibniz's Law is also known as the indiscernibility of identicals; it basically states that if two objects are actually identical, then they have all the same properties.
2. Of course, one is not required to be a full-blown nihilist or a full-blown universalist. One could believe that composition occurs in some cases, but not in others. Peter van Inwagen, for example, believes composition only occurs in the case where simples compose a living organism.
3. I take the term 'spooky' from Mark Balaguer in lectures on metaphysics (CSULA).
4. Thanks to Mark Balaguer for pointing out a possible uncharitable oversight.
5. See Bennett 2010, Section 8.
6. The idea that scientific theories are laden with metaphysical assumptions is due to lectures by Ricardo Gomez on the presence of values in science (CSULA).
7. See Bennett 2010, pp. 65-71.
8. See Bennett 2010, pp. 73-74.
9. One of the classifications of Thomasson as a nihilist is partially due to Mark Balaguer from lectures and discussions.

Bibliography

Bennett, Karen. (2009) "Composition, Colocation, and Metaontology," in: David John Chalmers, David Manley & Ryan Wasserman (Eds), *Metametaphysics: New Essays on the Foundations of Ontology*, pp. 38-76 (Oxford: Oxford University Press)

Chalmers, David J. (2009) "Ontological Anti-Realism," in: David John Chalmers, David Manley & Ryan Wasserman (Eds), *Metametaphysics: New Essays on the Foundations of Ontology*, pp. 77-129 (Oxford: Oxford University Press)

Sider, Theodore. (2009) "Ontological Realism," in: David John Chalmers, David Manley & Ryan Wasserman (Eds), *Metametaphysics: New Essays on the Foundations of Ontology*, pp. 384-423 (Oxford: Oxford University Press)

Thomasson, Amie L. (2009) "Answerable and Unanswerable Questions," in: David John Chalmers, David Manley & Ryan Wasserman (Eds), *Metametaphysics: New Essays on the Foundations of Ontology*, pp. 444-471 (Oxford: Oxford University Press)

The No Miracles Argument (NMA) for Scientific Realism: On Principle

Matthew Hart

> One thing I have learned in a long life: that all our science, measured against reality, is primitive and childlike—and yet it is the most precious thing we have.
> — Albert Einstein, letter to Hans Muehsam, 9 July 1951

Introduction

The NMA's tenability depends on one's definition of Scientific Realism. The central, unifying element present in the varying definitions of Scientific Realism (this element being the epistemically positive attitudes towards the theoretical claims derived from the results of scientific inquiry, whether the claims are about observable or unobservable entities) gains a robustness in light of the NMA that is difficult to discount. However, the varying definitions of Scientific Realism turn on differing views of the aim of science. Whether the NMA does or does not appear tenable depends on one's view of what its explanatory function is. Scientific Realism maintains that the NMA should be viewed as an axiomatic principle (e.g. like Occam's razor), which is essential to the practice of science. Opponents of the NMA argue that it goes beyond proper scientific explanation and question the reliability of the methods scientists use to arrive at their theoretical claims (Psillos 1999, pp. 71-76). However, a critique of the NMA also involves a critique of a form of reasoning called 'inference to the best explanation' (IBE). The inferential paths used in science are, in fact, an essential foundation to generating novel predictions and unifying disparate established claims. Therefore, the NMA is an axiomatic principle that underlies the practice of science and adds to the reliability of scientific predictions. The NMA remains a tenable account of the

philosophy of science.

I will first provide a short overview of Scientific Realism and the role of the NMA. Second, I will show how to formulate the NMA as an argument for scientific realism, showing that it is a form of IBE. I will then discuss IBE in general and its use in scientific practice. IBE is not used to defend Scientific Realism, but scientific methodology. The explanatory function of the NMA is to show the reliability of scientific methodology. We'll then look at objections to the NMA, starting with the possibility that the NMA is an instance of the base rate fallacy. I'll then discuss van Fraassen's Constructive Empiricism and show that it does not properly address the function of the NMA, nor does it provide a better account of why there are empirically and predictively successful scientific theories. Finally, I'll discuss Frost-Arnold's argument against the NMA and discuss how it misses the explanatory function of the NMA.

Scientific Realism

Part of Scientific Realism is the claim that the aim of science is to provide us with theories that—literally construed—are most likely to be true based upon the available evidence. There are varying, more subtle definitions of Scientific Realism, but for the purposes of this paper we'll work with this simple and general definition. There are two different aspects to this definition. The first one is semantic: it has to do with the language of science, the language in which our theories are formulated. Scientific Realism states that we should take the language of science at face value. If we have a theory, like atomic theory, which talks about electrons, we should understand the term 'electron' as referring to—or picking out—real objects in the world. This is true for all theoretical terms—the Scientific Realist is committed to saying that our theoretical terms refer to real objects in the external world. The second aspect is epistemic optimism. The Scientific Realist believes that whatever the theory says about objects (e.g. electrons) at least approximately corresponds to facts in the external world (e.g. the fact that

an atom is a basic unit of matter that consists of a dense nucleus surrounded by a cloud of negatively charged electrons), and believes that we have good grounds to think that our best theories are at least approximately true.

The No Miracles Argument (NMA) says that Scientific Realism is the only philosophy of science that does not make the success of science a miracle (insofar as the success of science would be a miracle if Scientific Realism is not true). The theoretical terms of our best scientific theories refer to real objects, those theories are (approximately) true, and the same term can refer to the same objects even if the term occurs in different scientific theories. These statements together are the best explanation of the incredible success of science. An actual account of how science develops, in the way in which Newtonian Mechanics replaced Aristotelian physics or Einstein's Relativity replaced Newtonian Mechanics, is that the older theory provides a partially correct and a partially incorrect account of a given object (or objects) that was replaced by a better account of the same object (or objects) in the subsequent theory. If the objects described didn't exist, if our theories were not really true—or at least approximately true—then it would be a miracle that our theories proved so successful in predicting phenomena. The NMA speaks to the reliability of scientific methodology, not to the reliability of IBE in general. The NMA is not a premise in defense of realism as such, but rather an axiomatic principle employed in scientific methodology.

The NMA is a form of IBE. We infer the hypothesis that would, if true, provide the best explanation of the available evidence. Thus, we infer the existence of extinct animals like Edmontonia, because this is the best explanation of the fossil evidence, just as we infer the existence of the Higgs Boson as the best explanation for the type of evidence that is produced by the Large Hadron Collider. We choose from a group of competing explanatory hypotheses—the one that we regard the best—the one that if true would provide a deeper understanding of the available evidence. The IBE is a powerful tool in the Scientific Realist's tool belt. The scheme of an IBE inference is as follows:

(1) A.

(2) B is the best explanation of A.

∴ B.

So, the argument for Scientific Realism can be shown as:

(1) Science is (empirically and predictively) successful.

(2) Scientific Realism is the best explanation of science's being successful.

∴ Scientific Realism.

It shows that the scientific hypotheses that we select and that we are willing to believe tend to be those that provide a deeper understanding and provide the best explanation of our evidence. We don't necessarily rely on our eyes or our instruments to believe in unobservable entities, but rather, we rely on the validity and robustness of our inferential practices—informed by a wealth of experimental data—to draw conclusions about what our universe is like. However, an IBE is not deductively valid: it is not the case that the premises logically entail the conclusion (it is not the case that the truth of the premises necessitates the truth of the conclusion). In fact, structurally an IBE is an instance of the fallacy of affirming the consequent:

(1) Y (phenomenon premise)

(2) $X \rightarrow Y$ [If X, then Y] (best explanation premise)

∴ X

The IBE can be made into a deductively valid argument by adding an inference-licensing premise as follows:

(1) Y

(2) X is the best explanation of Y

(3) [Y & (X is the best explanation of Y)] → X

∴ X

Now the burden of proof lies in justifying the inclusion of this premise. While the IBE may not be deductively valid, it is nonetheless a reliable inference. The argument is similar to the inductive vindication of induction; induction has worked in the past and it will continue to work in the future. The defense is circular but not viciously circular, because its circularity is not *premise* circular but *rule* circular. If the conclusion of an argument is also taken as one of the premises, then an argument is said to be viciously circular. Rule circularity is when the conclusion of an argument states that a particular rule is reliable, but that conclusion only follows from the premises when that very rule is used (Ladyman 2002, p. 218). Ladyman writes: "The conclusion that the use of IBE in science is reliable is not a premise of this defense of realism, but the use of IBE is required to reach this conclusion from the premises that IBE is part of scientific methodology and that scientific methodology is instrumentally reliable" (Ladyman 2002, p. 218). So, while the IBE is not deductively valid, it is at least as credible as inductive reasoning; and, as Lewis Carroll shows in "What the Tortoise Said to Achilles," even deductive reasoning cannot be defended by a non-circular argument. Carroll demonstrated that there are regress problems even in modus ponens deductions: in order to explain any logical principle, one must use a prior principle to justify the use of the logical principle.[1] However, if we take the logical principle to be axiomatic, then we can apply it *simpliciter*. Since science is the practice of making the universe comprehensible and miracles are (by definition) outside of the realm of explanation, science is in the business of dissolving miracles. Scientific methodology is founded on the principle that it is hitting empirical bedrock. In this light, the IBE is as axiomatically valid as inductive reasoning. Therefore, the IBE account of the NMA provides a tenable account of the NMA as an argument for Scientific Realism.

There are several criticisms of the NMA from the anti-realist perspective, such as the criticism that the NMA is an instance of the base rate fallacy or criticism of the idea that Scientific Realism is the best explanation of science's success. The base rate fallacy

occurs when one is judging the probability of an event and, having been presented with both generic information about the frequency of that event and specific information about the event in question, one tends to focus solely on the specific information and ignores the generic information or base rate. So, as it relates to the NMA, if the Scientific Realist is using the success of a specific scientific theory as a gauge of its approximate truth, and ignoring the base rate of all approximately true theories, then that could be considered an instance of the base rate fallacy. The success of a specific theory does not—in itself—imply that it is likely to be approximately true. Since there is no way of knowing what the base rate of approximately true theories is, the likelihood of its being approximately true cannot be determined. However, the Scientific Realist could respond by saying that the NMA is not stated in terms of probability. The NMA doesn't speak to the likelihood of our theories being true, it states that either our theories are true (or approximately true) or they are not true and it is a miracle that they have been so successful in predicting phenomena. A correct formulation of the NMA dissolves any appeal to the base rate fallacy.

van Fraassen's Constructive Empiricism

In this paper, we'll look at van Fraassen's Constructive Empiricism as an anti-realist response to the NMA. The Constructive Empiricist maintains that science aims at the truth about observable aspects of the world, but that science does not aim at the truth about unobservable aspects. Van Fraassen developed a Darwinian view of the survival of scientific theories:

> [T]he success of current scientific theories is no miracle. It is not even surprising to the scientific (Darwinist) mind. For any scientific theory is born into a life of fierce competition, a jungle red in tooth and claw. Only the successful theories survive the ones which in fact latched on to actual regularities in nature. (van Fraassen 1980, p. 40)

However, van Fraassen's explanation of the success of scientific theories merely explains *that* we have successful theories. It is one thing to explain why only successful theories survive; it is another thing to explain what makes a theory successful. The Scientific Realist would claim that he has *also* told a story about what makes a theory successful: the Scientific Realist claims that the theories that survive are successful because they are true, because the entities that they postulate are real, and because what the theories say about those entities is (approximately) true. The theories that didn't survive are the theories that failed, because they were simply false (e.g. the Ether theory, the Phlogiston, and the Caloric theory). So this is the first response to van Fraassen's reformulation of the NMA (Constructive Empiricism): Scientific Realism offers a superior explanation of the success of scientific theories since it explains both *why* and *that* scientific theories are successful, while Constructive Empiricism only explains *that* scientific theories are successful.

The second response attacks van Fraassen's distinction between observable phenomena and unobservable entities. Why should we not rely on our scientific instruments, whether they're electron microscopes or particle colliders, to deliver to us reliable images about what things there actually are in the universe? Why should we trust our naked eye more than our scientific instruments? The success of science at predicting new and surprising phenomena through the use of technology would be miraculous, the Scientific Realist would argue, if our theories were not correctly referring to the unobservable entities and their underlying processes that we can observe. Ladyman writes: "The coincidence in question would be that instruments and devices such as electron microscopes and microwave ovens mysteriously behave just like they would if there were atoms and electromagnetic waves" (Ladyman 2001, p. 213). The Scientific Realist thinks that we are justified in believing in unobservable entities because the inferential path that leads us to unobservable entities is the same inferential path that leads us to unobserved observables. No one has seen a dinosaur, yet if we were capable of traveling back through

time, we could, in principle, see a dinosaur with our naked eye. A dinosaur is a characteristic example of an unobserved observable. We are justified in believing in atoms, electrons, DNA sequences, and other unobserved entities on the same ground in which we are justified to believe in dinosaurs. Fossil evidence is what paleontologists use to reconstruct the history of our planet. From fossil evidence we can reconstruct important information about the life of extinct animals from the Cretaceous Period, like Edmontonia. We can come to know about their diets, the environmental conditions in which they lived, where they were geographically distributed, and more. But as fossils provide evidence for now extinct species, similarly—one could argue— the Large Hadron Collider's outcomes at CERN can provide evidence for the elusive Higgs Boson. The inferential path to the unobservable Higgs Boson is one and the same as the inferential path that leads us to the unobserved observable Edmontonia. Again, the NMA speaks to the reliability of the methods used in science, not specific explanations in a given theory.

FROST-ARNOLD

Frost-Arnold's (2007) critique of the NMA pertains to the aim of science. He claims that scientists (usually) do not accept explanations whose explanans-statements neither generate novel predictions nor unify apparently disparate established claims. Scientific Realism (as it appears in the NMA) is an explanans that makes no new predictions and fails to unify disparate established claims. Frost-Arnold writes: "Proponents of the NMA explicitly adopt a naturalism that forbids philosophy of science from using any methods not employed by science itself" (Frost-Arnold 2007, p. 35). Frost-Arnold's argument is given as follows:

(P1) Scientists do not accept explanans-statements that neither generate new predictions nor unify apparently disparate claims.

(P2) Scientific realism (as it appears in the NMA) is an

explanans-statement that neither generates new predictions nor unifies apparently disparate claims.

(P3) Naturalistic philosophers of science "should employ no methods other than those used by the scientists themselves" (Psillos 1999, p. 78).

∴ Naturalistic philosophers of science should not accept scientific realism (as it appears in the NMA).

Frost-Arnold writes, "the type of explanation that the NMA uses to explain the empirical success of science is exactly the kind of explanation… that scientists do not accept; that is, realism neither makes new predictions nor unifies previously disparate claims" (2007, p. 47). However, the truth of a specific scientific explanation (P1) and the function of the methodological principles (P2) that scientists use to arrive at that explanation are two distinct categories. Frost-Arnold's attack is directed toward scientific explanation and not scientific methodology. Frost-Arnold conflates scientific explanation and scientific methodology, and thus misses the aim of the NMA. The explanatory function of the NMA is to defend scientific methodology, not the truth of specific scientific explanations. A correct formulation of the NMA (as an axiomatic principle underlining the practice of science) dissolves Frost-Arnold's argument.

Conclusion

In this paper I provided an overview of Scientific Realism and the role of the NMA. I then showed how to best formulate the NMA and showed that it is a form of IBE. I discussed how IBE functions in general and specifically in the practice of science (i.e. IBE is not used to defend Scientific Realism per se, but the success and instrumental reliability of scientific methodology). I showed that a correct formulation of the NMA dissolves any appeal to the base rate fallacy. I then discussed van Fraassen's Constructive Empiricism as an example of an anti-realist critique of the NMA and demonstrated that it does not properly address the function of

the NMA nor provide a better account of why there are empirically and predicatively successful scientific theories. Finally, I showed that Frost-Arnold conflates scientific explanation and scientific methodology, and thus misses the aim of the NMA. A correct formulation of the NMA (as an axiomatic principle underlining the practice of science) dissolves Frost-Arnold's argument. Therefore, I have defended the view that the NMA is an axiomatic principle that underlies the practice of science and adds to the reliability of scientific predictions and remains a tenable view of the philosophy of science.

Notes

1. It is seemingly impossible for any method to justify itself. However, if we don't want to "saw off the branch" on which we're sitting, it appears that we have to posit axioms in any formal system.

Bibliography

Frost-Arnold, Greg. (2007) "The No-Miracles Argument for Realism: Inference to an Unacceptable Explanation." *Philosophy of Science*, 77.1, pp. 35-58.

Ladyman, James. (2002) *Understanding Philosophy of Science* (London/New York: Routledge)

Psillos, Stathis. (1999) *Scientific Realism* (New York: Routledge)

Van Fraassen, Bas C. (1980) *The Scientific Image* (Oxford/New York: Oxford University Press)

Scientific Metaphysics, Quantum Individuality, and Moderate Ontic Structural Realism

Ben Futernick

1. Introduction

Ontic structural realism (OSR) is a metaphysical position characterized by the thesis that objects are not ontologically 'prior' to, or more fundamental than, relations. Objecthood is difficult to define, but generally objects are just those things which stand in relations and are not relations themselves. OSR is generally argued for on the basis of its consistency with contemporary physics and its avoidance of problems plaguing other metaphysical views (e.g. odd metaphysical commitments). Different versions of OSR are demarcated primarily by varying versions of the metaphysical thesis. Eliminative OSR (EOSR) denies the existence of objects entirely, while non-eliminative OSR allows for the existence of objects and comes in two types, which I will call NOSR and moderate OSR (MOSR, named for Esfeld and Lam's position). NOSR holds that objects exist but that structure is 'ontologically prior' to objects, while MOSR holds that structure and objects are 'ontologically equal' (see Esfeld and Lam, 2008). Finally, I will call the set of positions which hold that objects are ontologically prior to structures 'object-centered realism' (OCR). This is the most common metaphysical thesis about the relation between objects and structures (particularly in modern philosophy) and is primarily characterized by a 'haecceitistic' theory of objecthood wherein objects are equipped with a 'primitive thisness' which individuates them from all other objects. Roughly, an object's haecceity is a property which is uniquely possessed by only that object; this is usually cashed out by defining an object's haecceity

as the property of being identical to that object. Further details of the haecceitistic theory are described in Section 5.

In this paper, I will first provide an account of the relevant quantum physics that is geared towards readers with only cursory knowledge of physics (Section 2). Then I will analyze prominent arguments for and against EOSR and NOSR, and ultimately reject both theories (Sections 3-4). I will discuss the tradition of OCR and put forth objections against haecceitistic theories, arguing that they do not provide the resources for the conceptions of objecthood and individuality required by contemporary quantum physics (Section 5). Finally, I argue (1) that MOSR is the best current metaphysical theory (in comparison to EOSR, NOSR, and OCR) in terms of internal consistency, consistency with contemporary physics, and achievement of the largest number of practical and theoretical desiderata of a metaphysical theory, and (2) that MOSR's status as the best current metaphysical theory provides sufficient grounds for provisionally accepting MOSR as literally true (Sections 6-7).

2. The Background Physics: Spatial Positions, Identity, and Exclusion

There are two types of fundamental particles in contemporary quantum mechanics: fermions and the bosons. Fermions are elementary or composite particles with half-integer spin, while bosons are elementary or composite particles with integer spin. The exact details of what spin is are irrelevant here; it just matters that fermions obey the 'Pauli Exclusion Principle' while bosons do not. The Pauli Exclusion Principle entails that two fermions of the same kind (two electrons, two neutrons, etc.) must have 'opposite' spin under certain circumstances, but same-kind bosons can share their spin. Particles can have either up-spin or down-spin, making spin a binary property. Further elements of spin and the Pauli Exclusion Principle will be explained when relevant later.

Broadly speaking, in the formalism of quantum mechanics, particles do not have definite spatial properties prior to measure-

ment. On the other hand, in classic mechanics, objects obey the 'Principle of Impenetrability,' meaning that two classical (i.e. non-quantum) objects cannot simultaneously occupy the same position in space. Additionally, in classical mechanics, objects not only have well-defined spatial properties (ex. trajectory and position among others) in the formalism but are also uncontroversially taken to *actually* have well-defined spatial properties. That is, if classical mechanics were true, then it would be uncontroversially true that no matter how difficult it is to find the position and momentum of a given classical object, that object *does* in fact have a well-defined position and momentum. On the contrary, there is contention in quantum mechanics over whether quantum particles *actually* have well-defined spatial properties prior to measurement. While most interpretations of (the formalism of) quantum mechanics tend to deny on technical grounds that particles actually have well-defined spatial properties prior to measurement, there are exceptions, most notably the De Broglie-Bohm interpretation, which I will return to later.

For now, I will bypass the question of whether quantum particles actually have well-defined positions and momentums prior to measurement; when I say particles share the same spatial properties (same position and momentum), I mean only that they are given the same *spatial wavefunction*, and vice versa. The details of this do not matter for this analysis, but roughly speaking, a particle's wavefunction gives the probabilities of states of that particle (i.e., the probability that a measurement of the particle will give a certain state rather than another). Therefore, a particle's spatial wavefunction gives the probabilities of that particle being measured as having certain spatial properties. In this paper, it will be relevant that there are cases of quantum particles sharing all of their properties. For simplicity, the case I will use in this paper is the case of a singlet state of two electrons in the first shell of a helium atom. Essentially, these two electrons have the same spatial wavefunction, and therefore have the same spatial properties; additionally, both being electrons, they have the same mass, charge, and so on. Finally, since electrons are fermions, they obey

the Pauli Exclusion Principle discussed earlier, and necessarily have opposite spins to each other. At first glance, then, we might think that these electrons *must* have different properties; after all, they have opposite spins! But prior to measurement, these particles *do not have definite spins*, in much the same way that they do not have definite spatial properties prior to measurement. Even though the two electrons do not have definite spins prior to measurement, thanks to the Pauli Exclusion Principle, we can still be certain that upon measurement these particles will have opposite spins; this will be important later. This has a bizarre result: these two distinct electrons are in fact completely indistinguishable. They share the same charge, mass, spin probability distributions, spatial properties, energy (because same orbital); there is no property that distinguishes them from each other prior to measurement. This is the foundation of Ladyman's metaphysical underdetermination argument, outlined in his 1998 paper introducing ontic structural realism, entitled "What is Structural Realism?"

3. EOSR: Metaphysical Underdetermination and Relations without Relata

According to the metaphysical underdetermination argument, contemporary quantum physics underdetermines whether quantum particles are individuals or non-individuals. An individual is (1) an *object* which (2) is *distinct* from every other object in some way, and (3) has an '*account of individuality*', that is, an *account* of what makes the object distinct from all other objects. For example, a baseball is an individual, and this baseball is distinct from every other object on at least two grounds. First, no two baseballs have literally every property in common: different baseballs are made of slightly different materials, or have tiny differences resulting from manufacture errors, or have impacted walls or mitts more frequently than others, or so on. Thus, we can be sure that the physical properties of any two baseballs will be different. Nonetheless, we can imagine two literally identical baseballs (same dents, same stitching, never been thrown, so on)

without any inconsistency, yet they are distinguishable from each other: they differ in their spatial properties. A baseball never occupies the exact same spatial location as another baseball, or rock, or bat; they are not the kind of things that can have identical spatial properties. So the 'account of individuality' (AOI) of a baseball could be a detailed account of its physical properties, excluding its spatial properties; a less restrictive AOI would include its spatial properties, distinguishing the identical baseballs in the thought experiment as well as the distinct baseballs in the real world example. There are various cases of quantum mechanical systems where neither type of AOI will count two particles as distinct. Two bosons with different spatial wave functions can nonetheless have the same charge, mass, energy, and spin (barring discussion of the indeterminacy of mass/charge/energy). So the first type of AOI would not individuate these distinct bosons. Perhaps the second, less restrictive AOI would individuate them; if the particles have distinct spatial wavelengths then the second AOI does in fact individuate them. But this will not always work. In the case provided at the end of Part 1, two electrons were shown to share all of their properties, spatial or otherwise, prior to measurement. Thus, neither type of AOI individuates these electrons; nonetheless, they are somehow distinct, since there are two electrons and not one. But all the properties of these electrons have been exhausted; all of their properties and relations have been exhausted and shown to be the same. Both electrons stand in the same spatial relations to everything else, both stand in the relation 'of opposite spin to' to each other, both stand in the same gravitational and electromagnetic relations, and so on; their mass, charge, energy (again barring discussion of indeterminacy) and spin probabilities are the same. What could there possibly be left to distinguish them from each other? It seems as though there is nothing left for physics to tell us about these particles.

Thus, Ladyman concludes that if "the quantum description is complete, then we are left with a dilemma: either PII is false, the quantum particles are individuals and there must be some principle of individuation of type (1) above; or quantum particles

are not individuals and PII is obviated in this context" (Ladyman, 2001 p. 65). The PII is Leibniz's Principle of the Identity of Indiscernibles, which states that any two objects that share all of their properties are identical, i.e. the same object. There are two forms of the PII, corresponding to our two types of AOI above; the first and stronger type of PII (PII-1) excludes spatial properties while the second and weaker version (PII-2) includes them. Thus, if there are two objects which share all of their properties, spatial and otherwise, and are nonetheless two distinct objects, then both forms of the PII must be false. If PII-2 is false, PII-1 must be false, since if two objects cannot be distinguished by their spatial properties nor their non-spatial properties, it is already known that they cannot be distinguished by their non-spatial properties alone (and therefore PII-1 is false). Then either (1) both forms of the PII are strictly false and quantum particles are distinguished by their haecceities or primitive thisnesses, or (2) at least the weaker form of the PII is true and quantum particles are in some sense not individuals. Option 2 could mean one of three things: (a) quantum particles are not objects; (b) quantum particles are not distinct from all other objects; or (c) there is no possible AOI which individuates quantum particles.

However, rather than arguing for option 1 or 2, Ladyman argues instead that we should drop our commitment to objects entirely, since their metaphysical status is underdetermined by the physics. According to Ladyman, we are committed to the existence of entities we do not know basic facts about, such as facts about whether or not they are distinct from other objects. Thus, he encourages us to shift to a new basis entirely: abandon objects and questions of individuality, and switch to 'structures', for which questions of individuality do not arise. Ladyman is an advocate of *eliminative ontic structural realism* (EOSR), a metaphysical theory that insists we abandon talk of objects from our metaphysical and scientific discourse and instead conceive of reality as being somehow structural at the fundamental level. I consider the argument that quantum mechanics underdetermines between haeccitistic OCR and non-individualistic OCR sound, and I will

return to it in Section 5. The most important point for the EOSRist to establish is the coherence of structures as an ontological basis.

The most common objection to EOSR is an objection to the possibility of structure as an ontological basis that replaces objects. This is the 'no relations without relata' objection. Since a structure is a set of relations holding between a set of objects, it seems that the idea of a structure without objects is incoherent; the existence of a structure requires the existence of relations, which requires things being related (i.e., things standing in relations to each other). This objection relies on a certain conception of what it means for one thing to be ontologically dependent on another thing. Ontological dependence is best defined as follows:

> Definition: A type of thing X is ontologically dependent on a type of thing Y if and only if (1) if some X exists, some Y must exist, and (2) if no Y exists, then no X can exist.

In the case of relations, it seems as though relations are ontologically dependent on relata (the things being related, the things standing in the relations). Take the relation "is larger than;" two things stand in this relation in virtue of one being larger than the other (this should be straightforward). Imagine someone telling you: X is larger than Y but neither X nor Y exists. This is an obviously false sentence. Things which do not exist do not have sizes, or any properties at all (ex. mass, charge, spin, position). This is not specific, then, to the "is larger than" relation, or any other relation; this is true of all relations. If the relata of a given relation do not exist, a statement of the type "X stands in this relation to Y" is false. It seems impossible for relations to exist without anything being related; additionally, the existence of a relation straightforwardly entails the existence of relata. We can conclude, then, that relations are ontologically dependent on relata.

EOSRists are therefore incorrect in believing that structures, as sets of relations without relata, can be a coherent ontological basis for scientific and metaphysical discourse. The ontological basis of a theory T is the most ontologically fundamental kind of thing which is part of that theory. Some type of thing Z is ontolog-

ically fundamental if and only if (1) it is not ontologically dependent on anything (except for something which is ontologically dependent on it) and (2) all other types of things are ontologically dependent on it. Since objects are *not* ontologically dependent on the relations they stand in but relations *are* ontologically dependent on objects (regardless of whether or not they are individuals), relations are not the type of thing which can be ontologically fundamental. Therefore, relations are an inadequate ontological basis for contemporary scientific and metaphysical theories, and EOSR is false.

As a note of intellectual honesty, this is not a completely decisive objection, though it is a very powerful one. It is reliant on a logical/set-theoretic conception of structure, but the EOSRist is welcome to redefine structure if they provide good reason for thinking it is superior to the logical conception of structure. For example, Jonathan Bain has attempted to bypass the 'no relata' objection by switching to a category-theoretic conception of structure. However, due to the limited scope of this work, I will not discuss attempts to rework the conception of structure; such attempts fall outside the scope of this paper and it is sufficient to show that EOSR is completely inconsistent with the most widespread conception of structure in metaphysical, scientific, mathematical, and logical discourse. On another note, I do characterize objects as ontologically dependent on relations later; this involves a complex reconstrual of the metaphysical nature of objects which is inconsistent with most views of objecthood. Therefore it would be more accurate to say: objects as generally conceived are not ontologically dependent on relations.

4. NOSR: OSR Beyond the 'No Relata' Objection

The instinctive response to the 'no relations without relata' objection is to conclude that for all relations there must be relata which stand in those relations. This response characterizes two variants of OSR: *non-eliminative OSR* (NOSR), a metaphysical theory

that holds that structure is in some sense 'ontologically prior to' objects, and *moderate OSR* (MOSR), which holds that structure and objects are ontologically dependent on each other; objects and structures are on equal 'ontological footing' so to speak. Ontological priority is defined as follows: X is ontologically *prior* to Y if and only if Y is ontologically *dependent* on X and not vice versa.[1] As a first point, the objection against the ontological fundamentality of structures developed in the last Section can be reapplied to NOSR here. Whether or not objects exist is not actually important to the argument; rather, the argument rehearsed in the last Section simply demonstrated that structures cannot be ontologically prior to objects, since they are ontologically dependent on them. So, although the NOSRist can arguably provide for the truth of sentences of the form "X is related to Y", the position is still inconsistent.

There is a second objection against NOSR, pioneered by Chakravartty (2012), which I will call the "exclusive disjunction dilemma." Chakravartty argues that the NOSRist must either (1) deny that objects have any intrinsic (i.e. non-relational) properties and therefore lapse into EOSR or (2) deny that structures are ontologically prior to objects (i.e. renounce NOSR). Therefore, according to Chakravartty, NOSR is impossible to maintain. The argument hinges on the idea that intrinsic properties guarantee non-relational identity. Chakravartty insists that "so long as the relata *have* genuinely intrinsic features… these intrinsic features keep popping up as possible candidates for determining their identity [the identity of the relata]" (Chakravartty, 2012 p. 197). However, even if we accept that *any* intrinsic property of a particle is a possible candidate for determining its individuality, particles have a finite number of properties. In the electron case we have been using so far, it is easy to demonstrate that the intrinsic properties of electrons are not candidates for determining their individuality. Electrons have four intrinsic properties: mass, charge, spin, and energy. It has already been shown that *none of these properties are candidates for individuality*. As for the exclusive disjunction dilemma, I showed in the last paragraph that one must

deny that structures are ontologically prior to objects, and just now I showed that there is no reason to think the NOSRist would have to deny that objects possess intrinsic properties anyway. The exclusive disjunction argument therefore does not gain ground against OSR.

NOSR, then, is an untenable position, but not for the reasons Chakravartty provides. Instead, it is untenable because it is impossible for relations to be ontologically prior to their relata. Relations are ontologically dependent on relata, as established in Section 3, so either relata are ontologically prior to relations, or relations and relata are on equal ontological grounds. This section is brief, but only because NOSR is disabled by the same objection rehearsed against EOSR; there is no need to overcomplicate its refutation.

5. Object-Centered Realism and Escaping Haecceitism

EOSR succumbs to the 'no relata' objection and the ontological priority objection, while NOSR succumbs to the ontological priority objection as well. At this point, there are two options available to the scientific metaphysician: a position that gives objects ontological priority over structure (OCR) or a position that places objects and structures on equal ontological footing (MOSR). In this Section, I will attempt to dispel the illusion that there are especially strong reasons to be an OCRist (an illusion that I believe to be the result of the historical dominance of OCR in philosophy and science, especially in the modern era).

OCR theories are either haeccitistic or non-individualistic.[2] Haecceitistic OCR (H-OCR) is characterized by two theses: (1) that every object is an individual, and (2) that these objects are individuals in virtue of possessing a unique property called a 'haecceity.' A haecceity is a *token property*, not a *type property*. Imagine the object-type 'bus'; there are many tokens of this type, namely, each bus. Each bus possesses a haecceity which is distinct from the haecceities of every other bus. In other words, each token of a type possesses a distinct haecceity, rather than buses in

general carrying a common haecceity. In H-OCR, the account of individuality for an object is as follows: X is distinct from Y, Z, etc. in virtue of uniquely possessing the property of X-ness (X's haecceity). A token object X's haecceity is the property of 'being identical to X'; Y's haecceity (Y-ness) is the property of 'being identical to Y'; and so on. The general purpose of haecceitism is to ensure that there are *two* particles rather than *one* in cases like the electron case where the particles are identical (disregarding haecceities). However, H-OCR fails to engage the problem of numerical diversity properly.

In the electron case, the two electrons share all their properties and, as a result, the question of whether they are identical is raised. The H-OCRist maintains that there are two electrons (X and Y) and posits that X has the property "being identical to X" while Y has the property "being identical to Y." The H-OCRist then asserts that since these two properties are distinct, X and Y are distinct. However, these properties would *not* be distinct if X and Y were identical: that is, the properties "being identical to X" and "being identical to Y" would be one and the same property. So the H-OCRist fails to show that there are two electrons: instead, the H-OCRist already assumes that there are two non-identical electrons. The NI-OCRist, on the other hand, can provide a solution which avoids excess metaphysical commitments and does not assume numerical diversity.

The NI-OCRist argues that these two electrons are numerically diverse because they stand in an *irreflexive relation to each other*. An irreflexive relation is, roughly speaking, a relation where something cannot stand in that relation to itself. For example, "is larger than" is an irreflexive relation; it makes no sense to say that one and the same thing is larger than itself (at a given moment). The two electrons we have been discussing must have opposite spins (by virtue of the Pauli Exclusion Principle) and therefore stand in the relation "is of opposite spin to" to each other. This is an irreflexive relation: nothing can have opposite spin to itself. The existence of an irreflexive relation guarantees the existence of two things, since there must be *two* things to stand in that rela-

tion. Hence, this relation between the two electrons guarantees their numerical diversity. NI-OCR, then, has the advantage over H-OCR in terms of metaphysical and epistemic parsimony (since it does not posit empirically inaccessible properties like haecceities) *and* provides a successful solution to the problem of numerical diversity. There seems to be little reason, then, to prefer H-OCR to NI-OCR; in addition, NI-OCR apparently has the resources to resolve the problems we have been discussing in a satisfactory way. If some form of OCR is sufficient to solve all these problems, then where did the fuss about structuralism come from? Why be a structuralist when OCR is sufficient? I will argue that NI-OCR and structuralism are closely connected: since NI-OCR provides an account of quantum individuality in terms of relations holding between particles, it is clear that some form of structuralism is already at play. In the next Section, I argue that NI-OCR collapses under scrutiny into a form of structuralism: moderate OSR.

6. NI-OCR AND MOSR

NI-OCR, in abandoning haecceities, has abandoned the ground for giving objects priority over structures without lapsing into eliminativism. Thus it collapses into MOSR, embracing the thesis that objects and relations are mutually ontologically dependent on each other. It has already been demonstrated that relations are ontologically dependent on relata,[3] but establishing the dependence of objects on relations is much more complex.

Objects are dependent on relations insofar as relata are not distinct entities without relations. In the electron case, each electron has certain intrinsic (non-relational) properties, such as mass and charge. In respect to these properties, the electrons are not distinct; in other words, numerical diversity is not achieved by reference to intrinsic properties alone. Even worse, the electrons actually share all of their relational properties as well. Both have the relational property 'is of opposite spin to the other electron', as well as sharing all spatial properties. If one of these electrons is considered on its own, there is nothing differentiating it from the

other electron considered on its own. However, these electrons become distinct, numerically diverse objects in virtue of being relata in an irreflexive relation. In other words, these two electrons are distinct entities solely in virtue of standing in certain relations; their metaphysical status as discrete entities is *structurally based*. So, in a frequent scenario, quantum objects are ontologically dependent on relations.

Then, on the quantum level, objects and relations stand in a relation of *ontological codependence*, characterized by two biconditionals, wherein: O = some object exists, and R = some relation exists. The biconditionals are:

(1) $O \leftrightarrow R$ [if some object exists, then some relation exists, and vice versa]

(2) $\sim O \leftrightarrow \sim R$ [if no object exists, then no relation exists, and vice versa]

Thus MOSR has been established: objects and relations are mutually ontologically dependent. In what follows, I provide a conjectural account of necessary and sufficient conditions[4] for the acceptance of a metaphysical theory, show that MOSR satisfies these conditions, and in doing so show that EOSR, NOSR, and H-OCR all fail to satisfy these conditions.

7. Moderate OSR and Good Reasons for Accepting Metaphysical Theories

An acceptable metaphysical theory must be (1) internally consistent and (2) consistent with our best scientific theories. If a metaphysical theory satisfies the above criteria and (3) has the greatest number of desirable traits in comparison to the other metaphysical theories that also satisfy those criteria, then we ought to accept it, since it is the best current metaphysical theory. In other words, the fulfillment of criteria 1, 2, and 3 is necessary and sufficient for the acceptance of a metaphysical theory. The most obvious objection would be that the criteria above may be necessary conditions for the acceptance of a metaphysical theory, but they are not suffi-

cient conditions. There are two forms of this objection: the anti-realist objection and the strict realist objection. The anti-realist wants to block the acceptance of metaphysical theories, while the strict realist wants to make it be 'stricter' about the acceptance of a metaphysical theory. The anti-realist argues that we have good reason to think there are no sufficient conditions for the acceptance of a metaphysical theory. The strict realist generally puts forth some kind of extra criteria which, when satisfied alongside the first three criteria, justifies the acceptance of a metaphysical theory. First, I will argue that the anti-realist has no principled grounds for holding such a position. Then I will argue that the strict realist objection may hold, but that this is not an issue.

The anti-realist objection is generally the insistence that there are no criteria such that their satisfaction is sufficient for acceptance of the satisfying metaphysical theory. Such a claim amounts to a sort of scientific quietism. EOSR, NOSR, H-OCR, MOSR, and eliminativism about relations altogether comprise an exhaustive list of theories of the relation between objects and relations. The anti-realist insists that we should accept none of them, regardless of their satisfaction of various desirable criteria, but this is an untenable position. Since this is an exhaustive list, if the conceptual framework of objects and relations is sound, one (and only one) of these must be true. So the anti-realist must either provide compelling reason to abandon the framework of objects and relations (a task which is daunting to say the least) or provide reason to think science should remain silent on at least some central questions about the ontological status of the fundamental things which make up the world. That is, either the anti-realist must argue for scientific quietism about the metaphysical status of objects and relations or argue for the abandonment of the overwhelmingly dominant metaphysical framework of science. Insofar as objects and relations are the subject matter of all natural scientific inquiry, staying silent on questions of whether they exist is undesirable and undermines scientific realism, so arguing for such quietism would be quite difficult. Since scientific literature and thought is deeply entrenched in the object-relation framework, it is *prima*

facie quite difficult to argue for the abandonment of such a framework. So the anti-realist objection is untenable.

There are two versions of the strict realist objection: a principled objection and an unprincipled objection. A *principled* strict realist objection puts forth new criteria, while an unprincipled strict realist objection rejects MOSR on basis of the possibility of new criteria. The *unprincipled* strict realist objection is that: since it is *possible* that the current account does not provide sufficient conditions, we do not have good reason to accept MOSR. In the absence of a good reason to think that the account is *actually incomplete*, as opposed to merely *possibly incomplete*, this objection is unpersuasive and intellectually lazy. If a principled strict realist objection holds against my account, that is not an issue; since this account is conjectural, I am fully open to the addition of new criteria if they are reasonable. I am of the opinion that the three criteria listed here are both necessary and sufficient for the acceptance of a metaphysical theory, but this is not strictly provable, since the acceptance of a metaphysical theory is inevitably based on certain theoretical values. So if a strict realist were to show me that: (1) the account is incomplete, (2) there are new criteria which are reasonable to add and provide sufficient conditions for the acceptance of a metaphysical theory, and (3) these criteria rule out the acceptance of MOSR, then I would relinquish MOSR. Having responded to the objections to my account of the conditions for the acceptance of a metaphysical theory, I will now show why MOSR should be accepted on this count.

MOSR avoids the 'no relata' objection, since it posits the existence of objects; it avoids the 'ontological priority' objection, since relations and relata occupy the same ontological level; it avoids the underdetermination problem, since it is non-individualist about objects; and it has various desirable theoretical traits, such as providing for numerical diversity in an elegant way and avoiding massive ontological overhauls (unlike EOSR), as well as various desirable practical traits, especially providing the resources for structural research programmes in physics. MOSR is internally consistent and consistent with the current state of

quantum physics (and general relativity; see Esfeld, 2006). Therefore, since MOSR satisfies all three conditions for the acceptance of a metaphysical theory, we should accept MOSR as a literally and approximately true description of reality.

Finally, MOSR is supported by an argument by elimination which has been generated over the course of this paper. Since objects must exist (since EOSR is false; Section 3) and relations must exist (since objects stand in spatial relations), both objects and relations exist. Relations cannot have priority over objects (since NOSR is false; Section 4) and objects cannot have priority over relations (since H-OCR is false and NI-OCR collapses into MOSR; Sections 5-6). Therefore, objects and relations must have equal ontological footing, by elimination. Objects can have equal ontological footing by either being ontologically independent of each other or ontologically codependent on each other. Since relations are ontologically dependent on objects and vice versa, objects and relations cannot be ontologically independent of each other. Therefore, objects and relations have equal ontological footing due to standing in a relation of ontological codependence (i.e., mutual ontological dependence).

8. Conclusion

In conclusion, contemporary physicists and philosophers of science should accept MOSR as being (approximately) literally true. Due to the scope of this essay, I was unable to discuss various things I would have liked to, such as MOSR in the context of general relativity and category-theoretic construals of structure as alternatives to the traditional set-theoretic/logical construal.[5] I would have also liked to discuss how ontic structural realism guarantees scientific realism; this is an underdiscussed (though not undiscussed) element of OSR which is absolutely crucial to its acceptance.

However, I would like to discuss one major note in the literature: Esfeld and Lam's version of MOSR. According to them, "Moderate structural realism proposes that there are objects, but

instead of being characterized by intrinsic properties, all there is to the basic physical objects are the relations in which they stand. Admitting objects provides for an empirical anchorage of the relations. Consequently, this position is not touched by the standard objection against the radical structural realism of French & Ladyman (2003) [the 'no relata' objection]" (Esfeld and Lam, 2008 p. 5). I rejected this casting of MOSR on the grounds that quantum objects do indeed seem to have intrinsic properties prior to measurement and the elimination of these is conceptually and practically difficult and unnecessary for the achievement of the desired theoretical results. In this way, my version of MOSR avoids another problem: in Esfeld and Lam's version, it is unclear what exactly is standing in the relations. If an object is nothing other than that thing which stands in relations, it seems that the objects are nothing in themselves, and therefore lapse into eliminativism.[6] My version of MOSR is simpler, more elegant, and achieves greater levels of practical and theoretical efficiency.

Notes

1. See p. 7 for the definition of ontological dependence.
2. See the discussion of underdetermination on pp. 6-7.
3. See Section 3, pp. 7-8.
4. If X is a *necessary* condition for Y, then Y cannot be the case unless X is the case; in other words, X is necessary for Y to be the case. If X is a *sufficient* condition for Y, then if X is the case, Y must be as well; in other words, X being the case is sufficient to guarantee that Y is the case.
5. See pp. 8-9 for discussion of the possibility of non-set-theoretic construals of structure.
6. See the discussion of Chakravartty's "exclusive disjunction dilemma" in Section 4.

Bibliography

Brading, Katherine and Skiles, Alexander (2012). "Underdetermination as a Path to Structural Realism". *Structural Realism*. Springer, 99-115.

Chakravartty, Anjan (2003). "The Structuralist Conception of Objects", *Philosophy of Science*, 70: 867-878.

Chakravartty, Anjan (2012). “Ontological Priority: The Conceptual Basis of Non-Eliminative, Ontic Structural Realism”. *Structure, Object, and Causality*. Ed. Landry and Rickles. Dordrecht: Springer. 187-206.

Esfeld, Michael (2004). “Quantum entanglement and a metaphysics of relations”, *Studies in History and Philosophy of Modern Physics*, 35: 601–617.

Esfeld, Michael (2009). “The Modal Nature of Structures in Ontic Structural Realism”, *International Studies in the Philosophy of Science*, 23: 179-194.

Esfeld, Michael and Lam, Vincent (2008). “Moderate structural realism about space-time”, *Synthese*, 160: 27-46.

Esfeld, Michael and Lam, Vincent (2012). “Ontic Structural Realism as a Metaphysics of Objects”. *Scientific Structuralism*. Springer. 143-161.

French, Steven and Redhead, Michael (1998). “Quantum Physics and the Identity of Indiscernibles”, *British Journal for the Philosophy of Science*, 39: 233-246.

Frigg, Roman and Votsis, Ioannis (2011). “Everything You Always Wanted to Know About Structural Realism but Were Afraid to Ask”, *European Journal for Philosophy of Science*, 1(2): 227–276.

Ladyman, James (1998). “What is Structural Realism?”, *Studies in the History and Philosophy of Science*, 29: 409-424.

Ladyman, James (2001). “Science, Metaphysics, and Structural Realism”, *Philosophica*, 67: 57-76.

Ladyman, James and Ross, Don (2007). *Every thing must go: Metaphysics naturalized*, Oxford: Oxford University Press.

Morganti, Matteo (2011). “Is There a Compelling Argument for Ontic Structural Realism?”, *Philosophy of Science*, 78: 1165-1176.

Psillos, Stathis (2006). “*The* Structure, the *Whole* Structure and Nothing *But* the Structure?”, *Philosophy of Science*, 73: 560–570.

Saunders, Simon (2006). “Are quantum particles objects?”, *Analysis*, 66: 52-63.

Existentialism and the Other: Reconsidering the Language of Simone de Beauvoir

Ashley Tarin

Introduction

The general idea of any work in the humanities is to call attention to current research in a given field. For many who plan to enter or are currently active in the academic profession, such a prescription might appear insignificant as it relates to the field of existentialism, a philosophy that has been deemed passé.[1] Existentialism continues to be taught as part of many philosophy curricula, but merely as a starting point to more prominent areas of study such as phenomenology and psychoanalysis.[2] The view that many professional philosophers hold is that while existentialism stands as a useful starting point to dabble in for undergraduates or non-philosophers, as its own discipline it is not an area for serious researchers in contemporary philosophy. Perhaps this is due to the cultural phenomena that arose from existentialism, namely the stereotype of the French intellectual wearing a black turtleneck, smoking a cigarette, standing in a dark alley, and reading a copy of Albert Camus' *The Outsider*.[3] I take the composition of this piece as a fortunate opportunity to counteract this all too prevalent perception. Rather than having a substantive argumentative basis, I take the devaluation of existentialism as a prejudice of historical custom. However, it is not so much that it has been precluded altogether, as it has been decomposed in order to fertilize various subsequent fields of thought.

Nonetheless, existentialism as a philosophy, as opposed to existentialism as a culture referred to above, is more evident in our everyday lives than a number of other philosophies, but

its influence has been rendered invisible. Existentialism represents one of the most significant recent attempts in philosophy to foreground the practical problems of living and relate them to the historical conditions of modernity. I will argue for this view through the exploration of the philosophy of Simone de Beauvoir; and, in particular, her conception of the Other and the notion of objectification.

The notion of the Other depicts prime examples of the aforementioned inevitable problems of living that are inherent in our world, but many turn a blind eye to. In an effort to draw attention to a specific structure at play between beings, my explanation of this notion will include a discussion of two of the senses in which the word *Other* is conceived: as a theory of groups and as our concern with how others define us. Having done this, I will discuss Beauvoir's notion of objectification, a notion that explicates the systemic existence of hierarchical relations between human beings. This notion rests on Beauvoir's distinction between subject and object. I will attempt to illuminate this distinction by looking at both ends of the spectrum: the process by which one can be objectified, as well as the occurrence of one objectifying others. At one end we have the process of objectifying others being regarded as "no big deal," and at the opposite end objectification is suddenly unacceptable when the tables are turned and we are the ones being objectified. Lastly, I will offer a look into the work of Judith Butler, specifically her conception of performativity, as it is articulated through and influenced by a reading and defense of Beauvoir. Ultimately, Butler reiterates Beauvoir's critique of a substantial notion of identity; in particular, the belief that identity is a matter of a fixed nature expressed through one's behavior. In agreement with Beauvoir, Butler's performativity articulates the idea that our behavior, instead of expressing our identities, constitutes our identities.

The First Sense: A Theory of Groups

The words *Other*, *Otherized*, *Othered*, etc., are prevalent in social contemporary criticism. For Beauvoir, to be an Other

is to be regarded as different from those against which you are being posited. As such, to be an Other is to be dissimilar and to be opposed to members of some common identity. The notion is concerned with how people come to think of themselves in relation to others, i.e. in opposition to others. For example, recall the days of attending high school. More than likely there existed cliques or groups of students that were different from one another: the popular kids, the nerds, the jocks, the cheerleaders, etc. Each group existed as the Other in relation to the rest simply because they were not the same. For Beauvoir, "the category of Other is as original as consciousness itself" (Beauvoir 2011, p. 6). This means that the process of Othering is a natural occurrence within human behavior. This act of categorizing and differentiating among beings should be regarded in the same way consciousness is.

This is to say that the first sense of the Other can be conceived as a theory by some group. According to Beauvoir, "no group ever defines itself as One without immediately setting up the Other opposite itself" (Ibid, p. 6). This means that no group of people can ever attempt to define themselves as a unified entity without an opposing group. So, we can say that the first sense in which the notion of the Other is conceived is more of a statement about the relations between groups of people. However, it is important to note is that the category of Other carries with it a more negative connotation. For instance, there exists a unified group (Group 1), and all outside groups (Group 2, Group 3, Group 4, etc.) are considered to be an Other in the sense that they are not the same as Group 1. The instance of not being the same is then identified as a lack of qualities, rather than a possession of merely different qualities. In this sense, qualities refer to genuine empirical differences. By lack of qualities I mean that there is a degree of fabrication that accompanies these empirical differences. For instance, a white man is empirically different to a man of color due to the shade of their skin, but to a white supremacist there is a degree of fabrication embedded in the quality of darker skin; namely, a sense of less-than-ness. Beauvoir mentions in *The Second Sex*,

> It only takes three travelers brought together by chance in the same train compartment for the rest of the travelers to become vaguely hostile 'others'… For the native of a country inhabitants of other countries are viewed as 'foreigners'; Jews are the 'others' for anti-Semites, blacks for racist Americans, indigenous people for colonists, proletarians for the propertied classes. (Ibid, p. 6)

Here we see the nature of duality and of opposing forces that inhabit man's ability to create understanding through relations in society, as well as the underlying negative nature of what it means to be an Other. These relations are systems that are not necessarily so much phenomena as they are central to and immediate givens in social reality. "This becomes clear when we consider consciousness that contains within itself a fundamental hostility to any other consciousness" (Ibid, p. 7), i.e. a subject posits itself in opposition and asserts itself as essential while setting the Other up as inessential. This theory of hostile consciousness is derived from the work of Georg Wilhelm Hegel, a philosopher whose work deeply influenced this concept of groups against groups that Beauvoir demonstrates. It will be useful to take a brief look into Hegel's work to better understand the prominence and commonality of Othering that, as a facet of existentialism, has been rendered irrelevant in modern times.

Hegel offers an account of other minds in his work *The Phenomenology of Mind*. However, his account approaches the existence of Others in terms of phenomenology and consciousness. Hegel holds the view that the phenomenal characteristics included in our experiences of Others are the only means we have of experiencing ourselves. In other words, we can only see ourselves when what we see is an additional self-consciousness, outside of our own. Knowledge of ourselves cannot be attained merely through introspection of our own dispositions and habits. We are not capable of discovering, let alone understanding, the concept of our own identity simply because we are not the only ones who exist in the world. We acquire the understanding of

"self-existence only in the self-existence of the other," and we must recognize ourselves as mutually recognizing one another (Hegel 1967).

Consider Hegel's example of Master and Slave, predominantly introduced into 20th century French philosophy by the lectures of Alexandre Kojève.[4] The Master and Slave relationship can be viewed much like Beauvoir's notion of an Other existing as an inferior in relation to a superior. In the first stage of the example, each being—both the Master and Slave—only recognizes their own conscious states and belittle any proposition of another existing consciousness. This, of course, risks the existence of their own self. However, Hegel explicates that to be called a person one must be willing to risk their own life in this manner. In both Hegel's and Beauvoir's philosophy, the issue is that this circumstance removes one of the necessary conditions; namely, that "whether one likes it or not, individuals and groups have no choice but to recognize the reciprocity of their relation" (Beauvoir 2011, p. 7). The Other must be acknowledged. Hegel then adds that in order to maintain a more sustainable theory of self-consciousness the relationship of one and the Other must be one of Master and Slave. To make it easier to grasp in the context of Beauvoir's work, rather than Master and Slave let us call them Man and Woman—i.e., woman being the inferior, the Other. Man exerts his so-called transcendence while denying it to woman and renders woman thing-like, i.e. an object. While man enjoys the fruits of woman's labor, he fails to realize that because of this inadequate relationship he cannot fully achieve perpetual transcendence or self-consciousness when woman, as the Other, to whom he must look in order to see himself, has been reduced to the status of slave.

It can be asserted now that women, for Beauvoir, are the Other to men. Beauvoir altered the original Hegelian context of the Other to explicate the presence, a presence that is equally prevalent now as it was then, of a male-dominated culture. This specific principle is the primary focus of her work *The Second Sex*. Perhaps one of her most widespread quotes, "One is not born,

but rather becomes, woman" (Ibid, p. 283), aims to rid the essentialism of women being born "feminine" and instead claims that they are constructed to be as such through social indoctrination. In this sense, to become a woman is to become an Other. Becoming a woman is to become something that is *not* a man, something that is inferior to man, something that is inessential. Beauvoir explores this view through the consideration of the education of woman from her childhood, through her adolescence, and through possible experiences of lesbianism and sexual initiation, should she undergo any at all. At each stage, Beauvoir illustrates how women are forced to relinquish their claims to transcendence and authentic subjectivity by a progressively more stringent acceptance of the "passive" and "alienated" role to man's "active" and "subjective" demands. Woman's passivity and alienation are then explored in light of woman's particular situations, such as being a wife, a mother, a prostitute, etc. Beauvoir studies these roles in life to show how women, instead of transcending through work and creativity, are forced into monotonous existences of having children, tending to the home, and being the sexual receptacles of the male libido. A more in-depth discussion of the foundations of this "woman as Other" school of thought, namely the underlying notion of objectification and distinction between subject and object, will be had in a proceeding section.

On the opposite end of the spectrum, as woman exists as the Other to man, man can exist as the Other for woman. However, this relationship must not be confused to be of the same nature as that of woman being the Other to man. When it comes to man as the Other, this placement only occurs in the sense that women naturally form a collective identity, a sisterhood, that stands in opposition to men. Just as the statement, "Woman is *not* man," was made previously to describe woman as the Other, i.e. inferior and inessential, the same statement can be made to describe man as the Other. Men can exist as the Other simply because they are not women. Even when men are placed as the Other, the negative connotation that comes with the category of woman seems to be precluded in the category of man. This is due to the fact that while

in some cases women regard men as the Other, women also have an acceptance of the idea, "I am not man, but I am here to serve him." She accepts this role as the Other for several reasons. One is that this nature of otherness appears to be absolute—"it lacks the contingent or incidental nature of historical facts" (Ibid, p. 7). Yes, woman regard themselves as *not* men in a sense that vilifies men, but this occurrence is not widespread and will never be simply because their relationships with men have become valued. If women could look back to a time when they were not inessential and see when and how they came to be such, they would know that it was possible, even predestined, that they not be the Other. A nature that was brought about by a certain set of conditions could be abolished by some other set of conditions. However, there is not such a time for them to look back to. Because of this, "she herself fails to bring about this change" (Ibid, p. 8).

Regarding this existential phenomenon, it is important to note that the focus should not be placed on whether or not people are being positioned explicitly as the Other, but rather the various forms this positioning takes. One does not need to name another self as the Other in order for this theory of opposing groups to exist. The mere mindsets of, "I am not them," "I am superior," "She is different than me," though do not appear necessarily as what people name "Othering," are in fact very common forms of the process that has always and will always take place in society. This principle demonstrates the assertion that the existence of Others is inevitable; it is universal, but it often characterizes itself as a notion that is not oppression.

The Second Sense: How Others Define Us

"Since it is the Other within us who is old, it is natural that the revelation of our age should come to us from outside—from others" (Beauvoir 1972, p. 288). In 1967, Beauvoir began a study of the same sense and caliber as *The Second Sex*. *The Coming of Age* critiques an additional prejudice held by society towards another oppressed group: the elderly. Though this second sense

of the Other might not be as easy to witness, it still takes the form of "Othering" in the sense that two groups, the young and old, are posited against one another, the former standing superior to the latter. Similar to the approach taken in the notion of woman as the Other to man, Beauvoir takes the fear of age as a cultural phenomenon as well, insofar as the elderly, as they stand in society, are a silenced and detested group. Again, this is an additional existential problem of living that has been unreasonably oversimplified. More work can be done here in the parallel case of elderly as Other, but for our purposes it is more useful to focus on her approach in *The Coming of Age* and what it says about our interpretation of what we are for ourselves and for others.

As she had done in *The Second Sex*, Beauvoir approaches the subject matter of *The Coming of Age* from a variety of perspectives; including the biological, anthropological, historical, and sociological. She points to an apparent gap between what we exist as for ourselves and how others interpret our existence. "Every human situation can be viewed from without-seen from the point of view of an outsider-or from within, insofar as the subject assumes and at the same time transcends it" (Beauvoir 1972, p. 10). Continuing to support her belief in the fundamental ambiguity of existence which involves the contradiction of immanence and transcendence, objectivity and subjectivity, Beauvoir in her approach on age attempts to capture what is not captured in terms of lived experience. The overall picture in context of this piece, nonetheless, is that how we interpret ourselves is contingent on how we interpret others interpreting us.

With this comes the notion of unrealizability, a concept Beauvoir presents that is evident in other philosophers' works such as that of Sartre, but has ultimately been influenced through her own writings. Primarily, it deals with our sense of what we are for ourselves and our sense of what we are for Others. According to existentialism, we can never exist as any one thing, i.e. we are never complete. Defining our essence is a process. The definition of who we are is not a fixed notion, once defined never to be changed. In due course, it is impossible to ever realize a closed

definition of who we are and it is our very state of being conscious that discloses this, meaning that so long as we are conscious living things, our essence will never be set in stone. Consequently, the definition of our essence is only fixed at death.

When Beauvoir claims that it is the Other within us who is old she is referring to the unrealizability of who we are. Specifically, how we interpret ourselves is not necessarily the way others interpret us. Likewise, the way others interpret us might not be the way we perceive ourselves or feel. Imagine an old woman. She appears to be elderly because you can see wrinkles in her hands and face, she is small in stature, she walks with a cane, etc. Thus, you define her as old. She herself, on the other hand, will never be capable of realizing that definition you have assigned to her because she feels she is the same person she always has been. She feels she has never changed. There exists an Other within her that is old, an Other who has aged, an Other who is closer to death. Think also about when you are given a compliment from someone. Although you enjoy hearing nice things being said about you, you never truly feel as if what the person says about you is true. There is always a strange awkwardness that accompanies perceiving the way others perceive you. This is because how we exist for them is not ever how we exist for ourselves. What and whom we exist as for ourselves is in constant conflict with what and who others expect us to be. For example, when we consider woman's place in society, for herself, her consciousness is a symbol of autonomy, an idea Beauvoir promotes. She might conceive of her own potential as limitless. Society, nonetheless, views her as a being on an opposite pole. There is an Other within woman who society sees as restricted, as inessential.

We see a similar existential notion developed in the work of Jean-Paul Sartre. In his work *Being and Nothingness*, Sartre introduces the concept of the Look. He includes it in his discussion of solipsism, i.e. the idea that I am the only conscious being in the world. His discussion of the Look has been referred to as a phenomenological proof of the existence of other people. However, for our purposes, we will focus on the involvement of judgment or

the process of "Othering" as apparent in the way others perceive us. Consider his example: You are looking through a keyhole, lost in observation, when you notice somebody behind you. Immediately you jerk back in shame, embarrassed to have been observed. The existence of the person who noticed you has intruded in your own self-existence. For Sartre, when we are looked at the world drains away from us towards the newcomer. The objects that only existed in relation to me before now exist in relation to this other person; she is judging them and observing them, just as I did, and worse still—she is observing me as well, as if I were part of the landscape. Our negative reaction to this experience is the feeling of shame. This is similar to Hegel's Master and Slave dialectic.

Unlike Hegel, however, Sartre does not think two people can look at each other in comfort and mutual recognition. For Sartre, it is always a contest: I look at you as an object (you are the Other to me), in response you turn it around and look at me (turning me into the Other), and the inescapable process continues. According to Sartre, human relations are defined by this conflict, similar to the view seen in Beauvoir's theory of groups as discussed in the preceding section.

The Notion of Objectification

Objectification is a notion closely related to Beauvoir's conception of the Other and is an important component in feminist theory. It is defined as the viewing of or the treating of a person, in this case the Other, Woman, or the elderly, as an object. Note that the terms *object* and *Other* will be used interchangeably in this section. While subjects, e.g. men, are said to possess the power to exist transcendently, i.e. the ability to be creative, active, and extend outward into the world, the Other exists only in a passive state. To be more explicit, man has denied woman the transcendent role. Women are forced to relinquish their existential right to transcendence and accept a circumscribed, repetitive imprisonment. There is no escape for them except through man, and even this is a dead-end. Man has projects, activities, accomplishments; woman just

has man. In this sense, objectification is a form of reduction, a state of less-than-ness. The subject, a transcendent being, belittles the object to immanence since the object is to always be overshadowed by the subject, who is essential and sovereign. Beauvoir writes:

> Indeed, beside every individual's claim to assert himself as subject—and ethical claim—lies the temptation to flee freedom and to make himself into a thing… the individual, passive, alienated, and lost, is prey to a foreign will, cut off from his transcendence, robbed of all worth. But it is an easy path: the anguish and stress of authentically assumed existence are thus avoided… Hence woman makes no claims for herself as subject because she lacks the concrete means, because she senses the necessary link connecting her to man… she often derives satisfaction from her role as Other. (Beauvoir 2011, p. 10)

We are offered here an image of the process by which woman, Other, has attained such status. An intuitive reaction might be one of suspicion in response to the fact that woman objectifies herself. Why would a woman want to be reduced to a passive state? According to Beauvoir, women acquire this sense of how to objectify themselves during adolescence. It is during this time in a girl's life that she accepts her femininity and engages in erotic transcendence. In other words, she realizes what it is to be wanted, to be an object of desire. This occurrence consists "in making herself prey in order to make a catch. She [willingly] becomes an object; and she grasps herself as an object…" (Ibid, p. 348).

Woman has learned to view herself from the outside in order to become something to be gazed at. This notion of consenting to objectification can be viewed as power or even, in a way, transferring the label of Other over to man. "Thus a vicious cycle is formed: less she exercises her freedom to understand, to grasp and discover the world around her, the less resources will she find within herself, the less will she dare to affirm herself as a subject" (Ibid, p. 308). Adolescence involves an existential crisis

for women in which they realize they must renounce any claim to being a Subject in order to be desired as an Object/Other. Beauvoir notes that it is partially through the act of looking that objectification occurs, pointing out that the male Subject's gaze "insensibly takes possession of the perceived image" (Ibid, p. 374). Same-sex female eroticism is thus reduced to a mutual narcissistic objectification, and/or as an attempt among others to reconcile her autonomy with the passivity of her flesh. Yet, though they struggle with the conflict of self-erasure to become the desired Object, young women are never truly able to achieve transcendence: "It is remarkable that in all those forms of behavior the young girl does not seek to transcend the natural and social order; she does not aim to extend the limits of the possible nor to work a transvaluation of values; she is content to display her revolt within the bounds of a world the frontiers and laws of which are preserved." (Ibid, p. 379)

Another form this notion of objectification takes is the form of myths or clichés. According to Beauvoir, myths are the labels assigned to women and can be viewed much like stereotypes. For instance, there is an insurmountable number of labels used to refer to women: mothers, wives, whores, virgins, emotional, irrational, etc. In *The Second Sex*, Beauvoir speaks of the particular myth of the eternal feminine, i.e. that vague so-called essence of femininity. While explaining that there is an overwhelming number of myths built up around the subject of woman, we must recognize their function as being devoted to sum her existence up in totality as a temporal being. In fact, however, our actual identities are always incomplete.

Beauvoir states, "In consequence, a number of incompatible myths exists, and men tarry musing before the strange incoherencies manifested by the idea of Femininity" (Ibid, p. 254). She uses this idea to support the claim, in both *The Second Sex* and *The Coming of Age* when she examines the myths of femininity and the myths of old age respectively, that while some may argue that a plurality of myths might not necessarily be a bad thing because it provides more "options;" and, it is never the case that one is able

to choose which myths are better suited for them as individuals. Even if one chooses to live up to the stringent social demands associated with a particular myth, the presence of various incompatible myths makes it inevitable that success in embodying one will lead to failure in embodying another. Thus, woman can never embody all of the qualities associated with each myth at any given point in time. I stress this claim to emphasize that myths are used against people, women and the elderly in this case, as means of setting them up for failure. The occurrence of continuously failing to become something acts as a suppressor to keep both women and the elderly in a state of passivity, rendering them inferior, thus maintaining the status of object or Other.

An Interpretation of Beauvoir as Articulated by Judith Butler

In regard to the relationship between identity and behavior, is identity something that is expressed through behavior or constituted by it? An argument can be made that it is a matter of a fixed nature. Our identity is given to us and is only expressed through acts of behavior. Influenced by studies in psychoanalysis, phenomenology, and speech act theory, whose ideas explore the ways that social reality is not a given but is continually created as an illusion "through language, gesture, and all manner of symbolic social sign", Butler offers a critique of this substantial notion of identity through her conception of performativity (Butler 1988, p. 519). The notion of performativity draws on the existentialist notion of identity in that it claims that identity is not a stable feature of our bodies from which various acts proceed. Instead, our actions only create the appearance of a fixed nature.

Taking recourse back to Beauvoir's claim that one is not born, but becomes woman, Beauvoir is asserting that gender as an identity is not a secure principle that beings are given and only after express through their actions. In defense of Beauvoir, Butler agrees that gender, and identity overall, is constituted over time through a repetition of acts. "Gender is instituted through

the stylization of the body and, hence, must be understood as the mundane way in which bodily gestures, movements, and enactments… constitute the illusion of an abiding gendered self (Ibid, p. 519). By the formulation of this idea, the conception of gender is altered from a substantial model of identity, i.e. from a static nature to one that involves the acceptance of an established social temporality. In other words, identity is constructed. It is an accumulation of performative acts that the audience, i.e. society, as well as the actor, come to believe.

A look at Sartre's waiter example will aid in the understanding of this concept.[5] There is a waiter who makes the greatest effort to conform to everything that a waiter should be. For Sartre, the waiter's exaggerated behavior is evidence that he is play-acting at being a waiter, a robot, so to speak, whose essence is to be a waiter. However, in order to play-act at being a waiter, the waiter must at some level be aware that he is not in fact a waiter, but a conscious human being who is deceiving himself that he is a waiter. This example shows that identity, ultimately, is given through modes of behavior to the extent that a being, the actor, becomes engulfed in the repetition of acts so much so that the act of being a waiter creates the appearance of a fixed identity.[6]

Where does the script for these apparent "actors" come from though? Historical conditions. Butler references Merleau-Ponty's reflections in his work *The Phenomenology of Perception* and states that the body is a historical idea. Likewise, in *The Second Sex*, Beauvoir sets the stage with the same reference, explicating that woman is a historical situation rather than a natural fact. Historical situation, in this sense, means that identity is performed by way of conventions set forth by history, i.e. past expectations, norms, beliefs, etc. Butler claims that the body, as a historical situation, endures an amount of cultural construction. This cultural construction she refers to occurs by means of "conventions that sanction and proscribe how one acts one's body, the 'act' or performance that one's body is, but also in the tacit conventions that structure the way the body is culturally perceived" (Ibid, p. 522). Here we see that the perception of one's identity is effected by

cultural perception insofar as the reproduction of identity occurs through the manner in which bodies are acted in relation to the embedded expectations of the outside world.

In addition to identity or gender being a performance in the ways discussed above, more significantly, it is a choice. We make reflective and pre-reflective choices. The former is the type of choice that we are readily aware of, such as buying a snack from the convenient store. The latter is the type of choice that we make "and only later realize we have made" (Butler 1986, p. 40). Sartre refers to this as "quasi knowledge," meaning that although it has access to our consciousness, we are not entirely conscious of it until after the fact. It is more of a spontaneous act, a pre-reflective choice. Beauvoir relies on this type of choice when referring to the kind of volitional act through which gender is assumed. Butler supports this idea as she explains in *Sex and Gender in Simone de Beauvoir's Second Sex* that taking on a gender or an identity is not something that occurs at the drop of a hat. It is a strategic process, a subtle process that only rarely is reflected upon and understood. Many people have the intuition to think "This is who I am, I choose who I want to be." Both Butler and Beauvoir anticipate this controversy because while it remains a choice, people are far less conscious of the specific choices they have made in developing their gender, let alone in developing their identity in general. Imagine waking up every morning and thinking thoughts such as "I am going wear a bow in my hair today because I wish to be perceived as a female," or "I will wear these pants more loosely because I hope that people will view me as a male." Thoughts like these do not occur simply because it is not reality. There is a societal and historical influence that plays a part in the choices we make that thrives in our subconscious until we are called to reflect upon it.

Conclusion

I have offered a discussion of elements in Simone de Beauvoir's philosophy that are only a few of existential insights pertaining to

basic—but often overlooked—features society. The notion of the Other is an attempt to illuminate the structure at play in the relations between human beings. This natural inclination to compare ourselves, as individuals and as groups, is a widespread occurrence that is intended to highlight specific features of inter-subjectivity that commonly go unnoticed. Furthermore, the attention given to the notion of unrealizability is utilized to explain the process by which we conceive of ourselves and how we are conceived of by others. A definition of who we are can never be reached because of this particular difference in perception. In addition, I have attempted to call attention to the notion of objectification as it relates to Beauvoir's conception of the Other, as well as how it plays a prominent role in feminist theory. To objectify a being, your own being or that of another, is to devalue them. This process explicates the existence of a systemic hierarchical relationship in which all beings participate. Lastly, I found it useful to take a look at the work of Judith Butler as it is understood through the philosophy of Beauvoir. When it comes to identity, both authors share the view that it is a matter of behavior constituting identity, i.e. identity is performed, and not a matter of behavior simply expressing a fixed nature already within us. Butler's notion of performativity shows apparent influence from the existentialist notion of identity.

It is without doubt that the previously discussed facets of existentialism are complicated characteristics of the past, present, and future relations between beings. Still, they continue to be oversimplified to the extent of losing value, especially in the philosophical world. While we can think of a list of many other areas of philosophy that, indeed, warrant the attention they have been given, existentialism should be reconsidered as an addition to that list, for the questions raised and conflicts apparent are ones which we face more often than people are aware of.

Notes

1. Existentialism is a term applied to the work of certain philosophers in the late 19th and 20th centuries. The philosophy asserted the belief that phil-

osophical thinking begins with the human subject, i.e. the feeling, acting, thinking, and living human individual. Existentialists share the belief that existence precedes essence. Some well-known existentialists include Jean-Paul Sartre, Soren Kierkegaard, Friedrich Nietzsche, Simone de Beauvoir, Martin Heidegger, et al.

2. Phenomenology is the study of the structures of experience and consciousness.
3. Albert Camus was a French novelist, journalist, and philosopher whose works depicted many existentialist themes and views. Although, Camus did not consider himself to be an existentialist despite the classifications of his works.
4. Alexandre Kojève was a philosopher responsible for the serious introduction of Hegel, specifically his Master-Slave dialectic, into 20th Century French philosophy, influencing many leading French intellectuals who attended his seminar on The Phenomenology of Spirit in Paris in the 1930s.
5. Jean-Paul Sartre elaborates on his example of the waiter in *Being and Nothingness*.
6. Special thanks to Professor Conway, a lecturer of Philosophy at Cal State University, Los Angeles, for helping me discover this connection between the notion of a fixed identity and Sartre's waiter.

Bibliography

Beauvoir, Simone de. (1972) *The Coming of Age* (New York: Putnam)

Beauvoir, Simone de. (2011) *The Second Sex* (New York: Vintage Books)

Butler, Judith. (1986) "Sex and Gender in Simone de Beauvoir's Second Sex," *Yale French Studies* 72, pp. 35-49

Butler, Judith. (1988) "Performative Acts and Gender Constitution: An Essay in Phenomenology and Feminist Theory," *Theatre Journal* 40, pp. 519-531

Hegel, Georg Wilhelm Friedrich. (1967) *The Phenomenology of Mind* (New York: Harper and Row)

Sartre, Jean-Paul. (1956) *Being and Nothingness* (New York: Pocket Books)

Compulsory Sizeism and Fat Existence as Resistance

Bree Lacey

> Nearly everyone wants to be normal. And who can blame them, if the alternative is being abnormal, or deviant, or not being one of the rest of us? Put in those terms, there doesn't seem to be a choice at all. Especially in America where normal outranks all other social aspirations.
> — Michael Warner, *The Trouble with Normal: Sex, Politics, and the Ethics of Queer Life*

The starting point for any analysis on compulsory systems is an interrogation of our notion of agency and choice. In Adrienne Rich's influential feminist essay "Compulsory Heterosexuality and Lesbian Existence," Rich analyzes how heterosexuality has been framed as an innate inclination that one would "naturally" choose (Rich 1980, p. 13). A byproduct of this framing of heterosexuality, as natural and as merely a choice, positions lesbianism as a divergence from one's natural inclination. Therefore, lesbian existence is "less" natural (Rich 1980, p. 26). Upon analysis of the various systemic and social forces that maintain and enforce heterosexuality such as the prohibition on women's sexuality; the force of male sexuality through rape, dowries, and economic dependency; the exploitation of women's labor by regarding women's labor as non-compensatory; denying them of a communal history; and by withholding them from pursuing education, it becomes clear that heterosexuality is more than just a natural inclination or mere choice, but rather is systemically and socially mandated through compulsory systems (Rich 1980, pp. 18-21).

In Robert McRuer's essay "Compulsory Able-Bodiedness and Queer/Disabled Existence," McRuer argues that the discursive and material moves of pushing non-normative or queer bodies

to the margins of society, to view their existence as mere choice, is found in societal relations of able-bodiedness and disability (McRuer 2002, p. 370). People with disabilities are defined by the negation of ability (McRuer 2002, p. 373). In a society governed by normativity and the desire to be normal, compulsory systems are guised as an exercise of agency and choice, when there is actually no choice at all. I argue that compulsory sizeism, or the discrimination of particular bodies on the basis of size, shares the compulsive features of compulsory heterosexuality and compulsory able-bodiedness. We are given the appearance of choice, but this illusion quickly dissipates when we look empirically at how those who are deemed non-normative or queer are treated, namely, as both "less" natural and as social deviants who need to be "cured." The hegemony of compulsory heterosexuality, compulsory able-bodiedness, and compulsory sizeism are constantly in danger of being disturbed and disrupted (McRuer 2012, p. 375).

The system of compulsory able-bodiedness is constantly demanding that those with disabilities personify the long-held belief that if they had the choice to become more able, that they would take it (McRuer 2012, p. 371). In a social landscape governed by compulsory able-bodiedness, compulsory heterosexuality, and compulsory sizeism, an appearance of choice is presented to those who are the outside of normative culture. Compulsory sizeism has to do with the society we inhabit rather than the particular fat bodies that occupy it. The social landscape we occupy is one that privileges thinness as inherently positive and something to strive for; and, the positioning of thinness as the "good" is achieved through normativity-enforcement and coercion, essential components of compulsory sizeism. When we look at how those who are deemed abnormal (i.e. those who do not subscribe to compulsory culture), they are perceived as deviant or failing to "be like the rest of us" (McRuer 2012, p. 370). In this way, compulsory able-bodiedness and, as a byproduct, compulsory sizeism are forms of normativity-enforcing coercion where one either subscribes or is socially othered and stigmatized.

I interpret both McRuer and Rich as arguing that compul-

sory able-bodiedness and compulsory heterosexuality are viewed as a kind of dogmatic, unquestioned, and unchallenged preference (Rich, p. 13; McRuer, p. 371). When we look at the phenomenon of sizeism, it appears that this pattern of assumed preference emerges. In a society governed by compulsory sizeism, one would never choose to be fat, in the same way no one would ever choose to be disabled or choose to be gay. Rich argues that two historical features of compulsory heterosexuality are physical force and consciousness-control (Rich 1980, p. 20). Following Rich, I argue that compulsory sizeism, although varied in application, contains the compulsory features of physical force and consciousness control that are found in her analysis of compulsory heterosexuality.

Disability Studies scholar Lennard Davis, in "Normality, Power, and Culture," details the ways in which the concept of the "normal" have led to the standardization of bodies via statistical data (Davis 2013, p. 3). Davis argues that one consequence of corporeal standardization is the societal division that occurs when particular bodies are labeled abnormal or nonstandard. A second consequence is that as a result of this division, the state will attempt to normalize those who have been deemed as nonstandard; a salient and primary feature of eugenics is the belief that the population can be normed and that corporeal diversity can eventually be ironed out. When we apply Davis's analysis of the standardization of bodies to compulsory sizeism, we see this normativity-enforcement at play. Within the medical industrial complex, fat people are often denied adequate resources for healthcare, given that obesity is viewed as a pre-condition for poor health and is, therefore, a liability for insurance companies (Berlant 2007, p. 763).[1] The New York Supreme Court shot down a law passed by Mayor Bloomberg that would limit the sizes of soft drinks being sold, just one legislative attempt to make fatness less visible, in the name of "health" promotion. I bring the soft drink legislation into my analysis, not because I, in some way, want Big Cola companies to continue profiting. I do so because it shows the way in which state, through legislation, seeks to normalize bodies under

the guise of healthcare. When we think of the standardization of bodies, in particular of body mass, the institutions that govern our lives can be seen as a means of exercising physical force through compulsory sizeism.

Consciousness control is an essential feature of compulsory sizeism. The phenomenon of compulsory sizeism is both internal and external. Due to the atmosphere of hyper-ability, hyper-productivity, and the pervasive yet elusive desire for the ideal body, everyone within a society governed by compulsory sizeism is encouraged to feel dissatisfied with their bodies. Even those who are the prototypical embodiment of health and fitness are always seeking to attain a better body. We inhabit a media-centric culture that relentlessly promotes and encourages weight loss. In *Fat in the Fire? Science, the News Media, and the "Obesity Epidemic,"* Saguy and Almeling argue that, although coverage of the "obesity epidemic" by the media has dramatized many facts about fatness, it also skews facts in order to highlight and increase individual blame of particular fat bodies that do not adhere to the compulsory sizeist ideology of chronic weight loss. In this way, compulsory sizeism exists both subjectively by internal policing of our bodies and objectively by policing the bodies of others.

The media is a salient tool in the production of compulsory sizeism. Although the media has never explicitly expressed its desire for fat non-existence, they have taken stealthy measures to mask the pervasive sizeist attitudes. Fat audiences are given particular "fat media tokens" as a kind of consolation prize to make visible the existence of fat people. Similar to McRuer, however, I argue that these fat media tokens seemingly render fat people intelligible, comprehensible, and visible while simultaneously obscuring and making elements of fat lived experiences invisible. These media tokens are supposed to communicate that fat people are no different than non-fat people; however, they obscure the oppression and sizeist discrimination many fat people face in their lives. Following the compulsory model, fat media tokens bring about confusing attitudes of how to perceive fat bodies in the media. For instance, we have particular fat women like Lena

Dunham and Gabourey Sidibe who are stars in their own rights, but are framed as such *despite* their being fat. These celebrities are often used as a means of showcasing the progress and diversity in the media by making fat people more intelligible, more digestible and as one could guess, these fat tokens are often the butt of the sizeist jokes that are so prevalent within the mainstream media. Another limitation of these fat tokens is that systemic structures that contribute to the production of fat bodies are rarely at the forefront of the discussion of these fat media tokens.[2] Particular parts of fat lived experience—namely, the intersection and contingency of where one is located within the social landscape as a fat person—is negated. These fat tokens make fat existence visible, but simultaneously distort it by negating the very real and felt intersections and complexities that make up the existence of fat bodies.

(Un)representation of fat people in the media via these fat tokens does not exist in a vacuum and it is important to analyze the context in which fat people are positioned. Fat celebrities are viewed as talented *despite* their fatness, and *despite* the societal pressure to conform to normative, corporeal ideals. Fat (un)representation in the media exists within a culture where there is a multi-billion dollar weight loss industry that constantly reinforces the ideological message transmitted via television shows, movies, and advertisements that says, "You're not attractive yet… but if you follow this diet, or go to this gym, or get this surgery, you could be." Susie Orbach, author of *Fat is a Feminist Issue*, states:

> Most people want to be slim, but this perceived physical perfection is difficult to hold on to and they fear losing control of it.... They project that fear and unhappiness on to people who are bigger and that often translates into abuse and attacks. It's a way of people disassociating themselves from what they fear the most—getting fat. (p. 34)

Fat bodies are only tolerated when they are on their way to thinness and, as a byproduct, more normal. We are constantly performing heterosexuality, able-bodiedness, and thinness (or the attainment thereof) but these performances are constantly failing.

Although the ideals are never attained, we continue these failing performances in hope that one day we will be able to achieve them (McRuer 2002, p. 372). A feature of consciousness control, under this system of compulsory sizeism, is the explicit mandate to view our bodies as liminal. Within our own psyches, we are made to position our bodies as moving from the unideal to the (unattainable) norm.

In her essay "Slow Death (Sovereignty, Obesity, Lateral Agency)," Lauren Berlant introduces the concept of slow death as "the physical wearing out of a population and the deterioration of people in that population that is very nearly a defining condition of their experience and historical existence" (p. 754). Berlant analyzes the obesity epidemic as the newest form of "slow death" and states, "health itself can then be seen as a side effect of successful normativity" (Berlant 2007, p. 765). Thus, those who are not deemed healthy are often perceived as abnormal; they are perceived as morally condemnable deviants who are unwilling to conform to the "natural" social order. Berlant states that epidemic crises put forth by the mainstream media are in no way neutral (Berlant 2007, p. 763). When the AIDS epidemic was at the forefront of media coverage in the 1980s, not only was there the salient health crisis at hand, the AIDS epidemic represented a moral crisis within American culture. Media coverage on the AIDS epidemic was framed as sexual deviants infiltrating and infecting normative culture. We can easily juxtapose the AIDS epidemic of the '80s with the contemporary obesity or "globesity" epidemic. Quickly, we see commonalities between the two (Berlant 2007, p. 758). If we look empirically at the obesity epidemic, similar to the AIDS epidemic, there is not only the initial crisis of fatness consuming endless amounts of government resources, there is the moral crisis of fat citizens who have infiltrated and are now infecting normative culture.

With the moral condemnation attached to fatness, we see a divergence between disability studies and fat studies. Whereas "natural" disability is viewed as an unfortunate, biological error, fat people are perceived as lacking discipline and self-respect, as

refusing to "take care of themselves;" and, therefore, they have an inability to contribute to the basic demands of capitalism (Berlant, p. 758; McRuer, p. 371). McRuer discusses the problematic elements of the definition of "able-bodied" put forth by the Oxford English Dictionary. To be "able-bodied" is to be "free from physical disability [which may potentially make it difficult to do the things that capitalism requires]". This definition incorporates the notion that those who are disabled are so in relation to the able-bodied person. It is argued that this is an essential characteristic of compulsory able-bodiedness. Those who are disabled are defined by the negation of ability. Furthermore, it is important to analyze the socio-political landscape that is found in Western culture where words like "disability" are utilized. Through a Marxian lens of capitalism, it is clear that able-bodiedness is not just to be free from disability, but also to be able to work. This definition of able-bodiedness benefits and maintains capitalistic ideologies where one's value is equated with their ability to produce. Philosopher Iris Marion Young in "The Five Faces of Oppression," defines marginalization as the act of moving particular oppressed groups from the productive core to the unproductive periphery thus potentially subjecting them to severe material deprivation or extermination (Young 2004, p. 53). This definition of able-bodiedness in conjunction with Young's conception of marginalization is useful when we apply it to the phenomenon of compulsory sizeism. If one does a quick Google search of "sizeism in the workplace," it becomes evident that instances of sizeism in the workplace have become increasingly more common. Sizeism appears to be not just limited to those on the far edges of the corporally abnormal periphery, but those who are components or potential components of the productive core. Along with the various forms of gendered violence women face in the workplace, fat women are at the receiving end of sizeist policing and appearance-based discrimination in the workplace because they do not occupy a normatively attractive body, which relegates them to the unnatural and queer. Those regarded as "too fat" or "too queer" are sidelined on the basis of potential healthcare liability and a lack of bodily

aesthetics, which renders them as a "poor" representation of the organization (Jones 2004, pp. 56-57). The intersection of gender and size serves as an instantiation of the interrelationship between compulsory able-bodiedness, compulsory sizeism, and compulsory heterosexuality.

In "Vulnerability and Resistance," Judith Butler calls out the neoliberal maneuver of framing the body in terms of individuality (Butler 2014, p. 11). We frame our bodies as singular and self-sufficient, which negates the systemic interactions and relational nature of our interactions. Butler posits that we view our bodies as dependent on systemic institutions and infrastructure while simultaneously viewing ourselves as dependent upon one another. She further argues that our binaural view of vulnerability, as opposed to agency and resistance, is an incorrect one. She argues that political vulnerability, or the subjection to institutional violence such as police brutality, can be a necessary condition for the mobilization of resistance (Butler 2014, p. 17). I reference the online community of "body positive" feminists who aim to eradicate fat-shaming practices both systemically and interpersonally (Null 2012).[3] In "Body Positivity: Identity, Community, and Self-Care on Tumblr.com," Kirschling provides an empirical analysis of the body positivity movement. Through a series of varying interviews with fat-identified individuals, nearly all of these women faced interpersonal violence through disparaging comments from observers that fatness should not be celebrated (Kirschling 2014, pp. 57-59). The severity of violence ranges from microaggressions that accuse fat activists of "glorifying obesity" to outright death threats (Kirschling 2014, p. 58).

In conjunction with Butler's analysis of political vulnerability as a means of resistance, body-positive activists are subjected to heinous acts of violence both systemically, through being barred from healthcare and being economically disadvantaged in relation to non-fat people, and interpersonally, through fat-stigma, body policing, and the imposition of unrealistic, normative beauty ideals. Compulsory sizeism is challenged by body positivity. In a society governed by compulsory sizeism, loving your body, as

it is, and not what it could be, is queer. If we can transcend the normative coercive measures used against us in order to maintain compulsory sizeist systems (and thus encourage us to normalize our bodies by attaining thinness), it becomes evident that our acknowledgment of our position within compulsory sizeism and our resistance to it serves as the very act of rebellion that corrodes this system of sizeism used against all of us who are governed by systems of compulsory sizeism.

Notes

1. My use of the term "obesity" is solely because I am talking about it in relation to the medical industrial complex. "Obesity" as a term for a disease says nothing about the "obese" body, only that it is fat and may potentially increase the likelihood of particular diseases.
2. My aim is not to vilify these women but rather to illuminate the process of systematic othering.
3. For more information on body positivity and the internet, see Alyssa M. Kirschling's "Body Positivity: Identity, Community, and Self-Care on Tumblr.com."

Bibliography

Almeling, Rene and Saguy, C. Abigail. (2008) "Fat in the Fire? Science, the News Media, and the 'Obesity Epidemic,'" *Sociological Forum* 23.1, pp. 53-83

Berlant, Lauren. (2007) "Slow Death (Sovereignty, Obesity, Lateral Agency)," *Critical Inquiry* 33, pp. 754-781

Butler, Judith. (2014) "Vulnerability and Resistance,"

Jones, Starlene. (2004) "Don't Call Me Skinny," *Off Our Backs* 34, 56-57

Kirshling, Alyssa M. (2014). "Body Positivity: Identity, Community, and Self-Care on Tumblr.com."

McRuer, Robert. (2002) "Compulsory Able-Bodiedness and Queer/Disabled Existence," in: L. Davis (Ed), *The Disability Studies Reader*, pp. 369-378 (New York: Routledge)

Murphy, Robert F. (2001) *The Body Silent: The Different World of the Disabled* (London: W.W. Norton Company, Inc.)

Null, Michaela Ann. (2012) "Fat stigma to fat acceptance: Fat women's size acceptance as an embodied process," (ProQuest, UMI Dissertations Publishing)

Orbach, Susan. (1997) *Fat Is a Feminist Issue* (United Kingdom: Random House Group)

Puar, Jasbir. (2012) "Coda: The Cost of Getting Better," *GLQ: A Journal of Lesbian and Gay Studies* 18.1, pp. 149-158

Rich, Adrienne. (1980) "Compulsory Heterosexuality and Lesbian Existence," *Journal of Women's History* 15.3, pp. 11-48

Young, Iris Marion. (2004) "Five Faces of Oppression," in: L. Heldke and P. O'Connor (Eds), *Oppression, Privilege, and Resistance*, pp. 39-65 (Boston: McGraw Hill)

Reluctant Kantian: Sartre's Misguided Rejection

John Dail

In his lecture *Existentialism is a Humanism*, Sartre makes several points, notably those which describe his conception of freedom, that seem to demonstrate some level of agreement with Kant's moral philosophy. However, Sartre explicitly rejects the Categorical Imperative, and thus Kant, based partly on a particular understanding of the Formula of Humanity formulation of the Categorical Imperative. While there has been some debate as to whether or not Sartre's use of Kantian language indicates an agreement with Kant's moral philosophy, I believe that such a debate is ultimately uninformative. Since Sartre continuously rejects Kant, any terminological similarities can be seen as merely coincidental and not an indication of broader agreement. Rather, it seems to me that Sartre's understanding of the Formula of Humanity is mistaken and, as such, his rejection of Kant is misguided. So, by first demonstrating this misunderstanding, and taking that as sufficient justification to reexamine Sartre's ethics, it is my intention to draw out some of the similarities between Sartre and Kant as more than mere coincidence.

I will begin my investigation by briefly presenting my method of investigation and explaining how it may lead to a positive result. Following the discussion of method, I will present Sartre's case of the "young man" which he believes provides a basic counter example to the Formula of Humanity because it results in a paradox. This seeming paradox will then lead to an in-depth procedural discussion of Kant's deduction of the Formula and the sorts of results that the Formula, when properly applied, will produce. With a better understanding of the proper application of the Formula, the discussion will return to Sartre's "young man"

in order to demonstrate that the case he provides does not actually result in a paradox. Finally, my investigation will conclude with a discussion of a few key points of Sartre's ethics, as presented in *Existentialism is a Humanism* and their possible connection to similar points in Kant's ethics. As a result of this discussion, I believe that I will be able to effectively demonstrate that there is good reason to believe that Sartre's ethics reflect a Kantian ethical picture that he would have otherwise denied based on a misunderstanding of Kant's work.

I. Critiquing the Method

Initially it seems that, when one attempts an investigation of similarities between two theories, a prudent method would be to make a one to one comparison of the uses of technical terms and their application within a theory. Such a method can tend to produce either one of two possible results: either the terms share the same or similar meanings in both theories and thus lead to similar conclusions, or the terms have distinct and different meanings in each theory and thus lead to different, and perhaps contradictory, conclusions. For example, if you and I were attempting to discuss how many units of distance it was from New York to Philadelphia, we would need to determine if we meant the same thing by the term 'unit of distance'. Without such a determination, it could very well be the case that we come to a disagreement about the distance between the two cities simply because I took a mile to be my standard unit of distance, while you were speaking in terms of kilometers. As such, we could both be speaking of units of distance and yet producing very different conclusions.

In one such analysis of the works of Sartre in comparison to Kant, Linsenbard points out that, "Sartre has been widely interpreted as invoking at least some elements of Kant's moral philosophy" (Linsenbard 2007, p. 65). If this is truly the case, then Sartre's explicit rejections of Kant seem particularly odd. Linsenbard continues to say that Sartre's obvious uses of Kantian terminology are such that, "he then radically modified [Kant's terms] for

his own purposes" (Linsenbard 2007, p. 65). The result of Sartre's modification is intended as a demonstration of his objection to Kant, meaning that he uses Kantian language against Kant by employing certain key terms in seemingly distinct ways. Further, even if it is the case that Sartre and Kant use the same terms with the same intention, Sartre's explicit rejections make such similarities seem merely coincidental in that they cannot be seen as indications of broader agreement. As such, beginning by looking for similarities can only result in the conclusion that Sartre and Kant stand only in opposition to one another since any seeming similarity is undermined by Sartre's explicit rejections. It is because of this that I believe to attempt a comparison by seeking out similarities in the uses of terms answers an uninformative question.

So, if it will ultimately prove to be uninformative to begin by seeking out similarities, how then ought one proceed? An alternative approach, suggested by Baiasu, is to look to the differences suggested by Sartre's criticisms, "what he assumes distinguishes him from [Kant]" (Baiasu 2003, p. 23). Baiasu explains that an investigation of this sort is less about assessing the success or failure of a particular objection to a position, but rather serves to open the door for a deeper investigation into whether or not the position was there to be objected to in the first place. If it is the case that the underlying position to which Sartre makes his objection is not actually Kant's position, then we are able to reject the rejection allowing for reexamination and assessment of any seeming similarities. Though this is a less conventional method of investigation, it avoids the problems that arise from seeking out similarity at the outset.

II. Sartre's Presentation of the Formula of Humanity

In *Existentialism is a Humanism*, Sartre is presenting a defense of a particular conception of existentialism by laying out, in a popularly accessible format, precisely what he takes existentialism to mean. As he points out, existentialist philosophy takes as a

starting point the premise that God does not exist. For many, this premise is extremely troubling because it gives reason to doubt that there can be such a thing as morality. Historically, moral law had been popularly viewed as the product of divine decree in the form of Holy Scriptures and commandments, so if these documents were falsified by the non-existence of God, then moral law would have no foundation. This simple deduction leads Sartre to discuss what he calls "abandonment." As explained by Sartre, "we are left alone and without excuse. This is what I mean when I say that man is condemned to be free: condemned because he did not create himself, yet nonetheless free, because once cast into the world, he is responsible for everything he does" (Sartre 2007, p. 29). It is this type of abandoned responsibility that Sartre intends to draw out as he presents the case of the young man.

The young man was a student of Sartre's who came to him seeking advice on how best to proceed in choosing how to handle a difficult dilemma. The conversation supposedly occurred during the German invasion of France in the time leading up to World War II. His father had left the family some time ago, and his brother had been killed in battle against the Germans. So only he and his mother remained of what had once been their family. The young man was presented with the opportunity to join forces fighting against the Germans, but this would mean leaving his mother alone. He believed, presumably with good reason, that if he left his mother to join the war, she would likely fall into great despair since he was her only remaining family. As well, he recognized the distinct possibility that if he chose to leave, his intention of doing well by fighting for a just cause could be thwarted for any number of reasons before he ever reached his destination. So his decision came down to a choice between a likely and tangible, albeit local, good, and an uncertain possible global good. Sartre sums up the dilemma by pointing to two distinct moral motivations: "a morality motivated by sympathy and individual devotion, and another morality with a broader scope, but less likely to be fruitful" (Sartre 2007, p. 31).

Sartre conjectures how the young man could decide. What

code could the young man follow in order to make the right decision? In seeking the answer, Sartre ponders a Kantian response in the form of the Formula of Humanity. According to Sartre, "Kantian morality instructs us to never treat another as a means, but always as an end" (Sartre 2007, p. 31). But, as he is quick to point out, this instruction fails to solve the problem. On the one hand, if the young man were to stay with his mother he treats her as an end. However, by staying home he treats those who fight the war as a means to his well-being, and by not joining and assisting them he fails to treat them as an end. On the other hand, if he leaves and joins the war he treats his fellow soldiers as an end by fighting alongside them, but treats his mother merely as a means by failing to act for her well-being. Should the young man choose either option, it seems he will inevitably be treating someone as merely a means and not as end. As such, it seems that Kantian morality, specifically the Formula of Humanity formulation of the Categorical Imperative, not only fails to solve the dilemma but also actually results in a paradox. So, Kantian morality must be false because if it were it true then the young man could simply plug in variables to the formula and the right action would be the result. So, Sartre rejects Kantian morality because it fails to produce answers to difficult questions.

This causes me to wonder, did the formula fail or did Sartre fail in applying the formula? When a mathematician produces an unexpected result does she blame the formula or does she check her work for error? It seems to me, as one who is prone to mathematical error, that it is at least possible that if the Formula of Humanity produces a paradox, certainly an unexpected result, then it is proper to investigate the application of the Formula. If it was properly applied, then Sartre is obviously onto something and we should reject Kantian ethics as, at least, problematic or unhelpful, if not altogether false. However, if there was an error in Sartre's work, then we owe it to Kant to fix the error and reapply the formula to find if it produces a useful result.

III. Kant's Formula of Humanity

In performing the due diligence of checking Sartre's application of the Formula of Humanity it is only proper to turn directly to Kant's *Groundwork of the Metaphysics of Morals*, wherein he provides the deduction of the formulations of the Categorical Imperative and demonstrates their application. Kant presents the Formula of Humanity as saying, "so act that you use humanity, in your own person as well as in the person of any other, always at the same time as an end, never merely as a means" (Kant 2012, p. 41). This is clearly a different phrasing than was used by Sartre, but is the difference in wording substantive, or just a matter of paraphrasing on Sartre's part? I contend that it does constitute a significant difference and that it does affect Sartre's understanding of the Formula. As such, it is prudent to understand precisely what Kant means by the words he uses and how he arrived at this formulation.

According to Kant, all things in nature act according to laws or principles. By this he means that all things in nature are restricted by certain physical laws and conditions that make it the case that particular things function in particular ways. For example, there are facts of the matter that make it the case that a bird can fly and a human cannot, or that an oak tree can produce an acorn while a human produces live human offspring. These types of facts of the matter can be seen to accord with laws that dictate what a given thing can do, and give us some ground for determining whether or not a thing is a good instance of its kind. For example, it would be a mistake to judge the goodness of a tomato on some quality that it is outside of its nature to have. We could judge a tomato based on ripeness or taste or color, but it would be ridiculous to hold a tomato to the standard of a piece of construction material. It is not in the nature of the tomato to be the sort of thing that we could use to construct a house. Similarly, we ought to judge human activity based on the nature of its kind. This brings us to the question, what is the nature of human activity?

Human beings have demonstrated themselves to be the sorts

of things that set ends, meaning they choose what they want and this choice determines them to action. But it is not simply a matter of saying, 'I want (x), so I will have (x).' Humans are equipped with and employ various aspects of the faculty of reason to figure out how best to proceed. In terms of action, humans employ practical reason; that aspect of human reason that applies to the practical world. So when a human wills herself to achieve some end, the will just is practical reason. Practical reason makes use of principles in determining how best to proceed in order to achieve an end, because reason tells us that it is in the nature of things to act in accordance with laws or principles. But how do we know if the will is good?

Since we have seen that the only proper way to judge something as good is to determine if it is a good instance of its kind, it follows that the only way to judge practical reason is to determine if it is a good instance of practical reason. So, in order for an instance of practical reason to be a good instance of its kind we must look to the principles, or laws, of practical reason to find the ground for judgment. Since practical reason is that aspect of reason that determines how best to proceed in order to achieve some end, practical reason also tells us that the best way to proceed in judging an instance of practical reason is to judge it according to its principles or laws. So, if the will just is practical reason, the will is good if, and only if, it is good with respect to the principles of practical reason. The instance of practical reason, then, can only proceed if it makes use of true principles, because if the principles it used were false, it would create a contradiction and practical reason ought to reject it. It follows from this procedure that the genuine products of practical reason are good as such. The engagement of practical reason in this manner reflects Kant's Formula of Universal Law of the Categorical Imperative. This does not, yet, get us to the Formula of Humanity, but it is a demonstration of practical reason's commitment to prior principles.

In order to continue the progression to the Formula of Humanity, it is necessary to understand more precisely what Kant means by the term "humanity" since the Formula directs us as

to how we ought to "use the humanity" in ourselves and others. From this wording it becomes clear that what Kant is referring to is some aspect of a person, rather than that person as such. Since it is necessary for the Formula to be making use of some objectively true principle, it must be some objectively true principle about people. It must be connected to or derived from the nature of humans. We have seen that human beings are end setters and that they make use of practical reason to determine the best course of action in setting out to achieve those ends. So it follows that it is an objectively true principle that humans use practical reason, and further that we conceive of others and ourselves as practical reasoners. As such, it follows that we must accept the principle of practical reason as the true principle of goodness. So, when Kant refers to "using the humanity…" he is actually referring to the principle of humanity, and the principle of humanity just is the principle of practical reason. As a result, the Formula of Humanity requires that we not use the principle of practical reason in others or ourselves merely as a means but must also set it as an end in itself. This idea can be fleshed out if we take a look at some of the examples Kant uses to demonstrate the application of the Formula, and the results it will produce.

A first example is that of making a false promise, or, more broadly, lying. Suppose I set as my end getting some significant sum of money from you, and you will only agree to loan me the money on the condition that I pay it back at some prearranged time. Based on that condition, I know that you would set it as an end that you recoup the money you've loaned. I understand the fact that if I make the promise, then you will give me the money. As well, I understand that if I don't make the promise, then you won't give me the money. I use this understanding about you to make the false promise that I will repay you at the prearranged time. By using my knowledge in such a way as this, I represent my end as good and your end as bad. To represent ends as either good or bad is akin to seeing an end as either worthy or unworthy of fulfillment. However, if both ends are the product of practical reason, and good as such, I must be able to represent them both as

good. By denying the goodness of the output of practical reason, I am denying the goodness of practical reason. It is in this way that I violate the Formula of Humanity, by using the principle of humanity merely as a means to my end of getting the money from you. By employing practical reason to achieve my end, I am committed to a prior principle that obligates me to maintain the goodness of practical reason as such.

In another example of proper application of the Formula, consider a case from beneficence. Suppose I am generally well off, and not in need of assistance. As Kant says, "humanity could indeed subsist if no one contributed anything to the happiness of others while not intentionally detracting anything from it; but this is still only a negative and not positive agreement with *humanity, as an end in itself*, if everyone does not also try, as far as he can, to advance the ends of others" (Kant 2012, p. 42). By this, Kant means that we have an imperfect duty to act for the benefit of others. It is not enough to simply avoid preventing others from achieving ends, but insofar as we are able, the Formula of Humanity requires us to take on the ends of others and make them our own. It is evident from the nature of practical reason and the goodness of its products, that we must represent the ends of others as good.

At this point it seems possible that one might object by claiming that one could set the end of killing another, and that based on the presentation of the Formula thus far we must see that end as good. However, this objection fails to take into account that the end of killing another fails to accord with the Formula of Humanity because in order for that end to be good, the end of the person to be killed, presumably to continue living, must be represented by the would-be killer as bad. We must bear in mind that the difference occurs at the level of practical reason, and that the instance of practical reason is what is truly subject to judgment. In this way, we can see that ends, meaning the products of good instances of practical reason, are good as such, while failed ends, meaning the products of instances of practical reason that do not make use of a principle, are bad.

IV. Reapplying the Formula

With this understanding of Kant's Formula of Humanity at hand, we can return to Sartre and evaluate his application of the Formula. Recall that Sartre interprets the Formula as instructing us not to treat another person as a means but rather as end. From what I have just presented, it becomes evident that Sartre's "Golden Rule" style interpretation is not inline with Kant's actual formulation. This demonstrates that, on an important level, Sartre has either critically misunderstood Kant or is purposefully misrepresenting his work. It seems to me to be more charitable to presume that Sartre would not intentionally misrepresent the work of another, so I will make the assumption that a misunderstanding is more likely and proceed from there. As a result of this misunderstanding, Sartre is doomed to misapply the Formula of Humanity, and we should not be surprised that his application results in a paradox. Similarly if I misunderstood geometry, it would not be surprising if I misapplied the Pythagorean Theorem and failed to calculate the proper length of the side of a triangle. So, it is now appropriate to revisit Sartre's case of the young man and see if we can produce a result for him that could solve his dilemma and guide him in action.

Given that the young man is attempting to choose between going to war to assist in that effort, and staying home with his mother because he is concerned for her welfare, we must begin by understanding what principles are at play. Clearly, and most obviously, the principle of humanity is part of the equation, but in order for either of these two options to truly be options for the young man he must be committed to some other narrow principles that tell him that either of these options are good or worthy to begin with. For example, in order for me to struggle with deciding between steak and salad while looking at a menu in a restaurant, I must already take it as good to eat meat. If I were committed to a prior principle that instructed me that eating meat was wrong, I would not even consider steak to be an option for dinner.

So, consider the young man's options. To stay at home with his mother would require commitment to a prior principle

of care or compassion. To go to war means that he is committed to some principle that tells him that fighting in a war, at least under a certain set of circumstances, is good. But if we return to Sartre's text, we find the young man's actual reason for wanting to go to war. Sartre states, "his older brother had been killed in the German offensive of 1940, and this young man, with primitive but noble feelings, wanted to avenge him" (Sartre 2007, p. 30). So the young man is actually committed, not to some prior principle about fighting the good fight, but rather he demonstrates a commitment to a principle of vengeance. So, in effect, what the young man is deciding between is a principle of vengeance and a principle of compassion. When we consider what the Formula might tell us with regard to these options, the result ought not to be surprising. To hold vengeance as an end is to take as a principle that it is good to inflict the same lack of regard for the ends of another as was inflicted on the person whom one seeks to avenge. This is clearly a violation of the Formula because it uses the principle of humanity, or rather a prior instance of the denial of the principle of humanity as the ground for a further denial of the principle. That is to say, because you denied the humanity in my brother it is right that I deny your humanity in return. Based on the result of applying the Formula to this option it seems evident that the right choice for the young man is to stay at home with his mother, but for the sake of completion we should run that option through the Formula as well.

For the young man to hold compassion as an end is to take as a principle that it is good to care for one who is need of caring. Thinking back to the case from beneficence, we can recall that the Formula results in an imperfect duty to take on the ends of others as our own, meaning that we ought to adopt the ends of others insofar as we are able. So the Formula commands that the young man take his mother's end of well-being on as his own, but not so much so that it prevents him from setting his own ends. As such, it is the young man's duty to care for his mother insofar as he able while still maintaining his own ends with regard to his own self-perfection.

As we can see, had Sartre had a better understanding of Kant's Formula of Humanity he would have been able to provide the young man with more detailed assistance than to say, "you are free, so choose; in other words invent" (Sartre 2007, p. 33). He could have told the young man that his desire for vengeance is the product of a bad instance of practical reason, one that could only result in contradiction. Further, he could tell the young man that he ought to stay at home with his mother out of compassion for her welfare, but not to take that as reason to sacrifice his own ends because he is, after all, free.

V. Sartre as Kantian

Now that we have seen that, with a proper understanding of the Formula of Humanity, Sartre could have provided the young man with guidance in solving his dilemma, it is justifiable to reject Sartre's rejection of Kant. Further, since we are able to set aside the rejection, it may now be fruitful to investigate Sartre's work to find places where he might agree with Kant.

For instance, the previous section closed with the word "free." Sartre uses the term "free" in a particular way so as to ascribe an ultimate responsibility to individuals. For Sartre, to be free makes it the case that we are responsible for ourselves; that we must be in control of our passions and not the other way around. This notion of being free mirrors Kant's notion that we are free insofar as we obligate ourselves to action. Both philosophers seem to share the view that to be free serves, in a sense, as a restriction. For Sartre, being free forces us to choose our actions in such a way that we cannot but be responsible for own our choices and the situations that follow from that choice. Likewise, for Kant our being free can only be demonstrated through the obligation of ourselves to unconditional practical laws that we give ourselves as the result of applying reason. Our freedom makes us accountable and responsible, and empowers us to restrict ourselves from acting solely in the interest of satisfying base desires.

Both Sartre and Kant also come to similar, though differently

worded, results of failing to accept or act from freedom. Sartre, in typical fashion, uses harsh language to name those who deny their freedom, calling them "cowards" or "bastards." The crime these people commit by denying their freedom is to live in "bad faith." Simply put, one who is in "bad faith" is one who denies their situation, that is to say, one who denies one's own reality inclusive of all of the circumstances of one's existence. These circumstances can include mental faculties and physical abilities, as well as any social forces at play. In terms of freedom, a person guilty of "bad faith" is someone who tries to escape responsibility for his action by claiming that he was determined to act in such a way so as to perform that action. In this way to be in "bad faith" is to freely choose to be enslaved by the passions, by desire, or by society.

Similarly, Kant would say that to deny one's freedom is to deny the principle of humanity. We can see that we are free because we set ends and set out on a course to achieve those ends. In setting a course to achieve ends we employ practical reason to determine the best course of action. It is through restricting ourselves to the best course of action that we demonstrate our freedom from mere desire; we are able to overcome it and be moral. It appears then that both Sartre and Kant have similar, though differently named, ideas regarding what it is to deny the sort of thing that we are as humans, and in both cases the result is to be immoral.

Again, however, during his discussion of freedom, Sartre makes the move to reject Kant specifically by claiming, "Kant states that freedom wills itself and the freedom of others. Agreed. But he believes that the formal and the universal are adequate to constitute a morality. We, to the contrary, believe that principles that are too abstract fail to define action" (Sartre 2007, p. 49). He backs up this rejection by referring us to the case of the young man, and how Kant's Formula fails to produce an adequate result. But as we have seen, if the Formula of Humanity is properly understood and applied, Kant can solve the young man's dilemma, so Sartre's rejection fails. Sartre does not follow the "abstract" principle to its concrete application. So, again, we don't need to take Sartre's objection as reason to prevent an investigation into simi-

larities between him and Kant.

VI. Conclusion

In closing, it appears that there is at least some similarity between these two great thinkers and that such similarity can be seen as more significant than mere coincidence, perhaps even as an indication of broader agreement. The significance of these similarities comes into view once we are able to move past Sartre's explicit rejection of Kant. Even though I have only scratched the surface of similarity in the previous section, I contend that if Sartre derives his ethics from his conception of freedom and being free, then it is possible that more subtle similarities to Kant's ethics can be found. Does this mean that Sartre is some sort of reluctant Kantian? Is he living in bad faith by denying his underlying agreement with Kant? Probably not, but it is nonetheless interesting that he so loudly rejects Kant based, at least partly, on a misunderstanding of the formulations of the Categorical Imperative, particularly the Formula of Humanity. Perhaps, at the very least, if Sartre had come to a proper understanding of the Categorical Imperative and its concrete applications he would not have been so quick to dismiss Kant. Certainly more work needs to be done in this area in order to arrive at a more definitive conclusion, as I have restricted myself to only a single, relatively short, informal work of Sartre's. So, by enabling ourselves to move past misguided rejections, we have opened the door to further investigations of underlying agreement, which could enhance our understandings of both Sartre and Kant.[2]

Notes

1. The emphasis in the quote is Kant's.
2. It is prudent that I mention that I have arrived at this interpretation of Kant's Formula as the result of a class on Kant's ethics, which I attended while an undergraduate at UCLA, taught by Ira Richardson (PHIL C115, Spring 2014).

Bibliography

Baiasu, Sorin. (2003) "The Anxiety of Influence: Sartre's Search for an Ethics and Kant's Moral Theory," *Sartre Studies International* 9(1), pp. 21-53

Kant, Immanuel. (2012) *Groundwork of the Metaphysics of Morals* (New York: Cambridge University Press)

Linsenbard, Gail. (2007) "Sartre's Criticisms of Kant's Moral Philosophy," *Sartre Studies International* 13(2), pp. 65-85

Sartre, Jean-Paul. (2007) *Existentialism is a Humanism* (New Haven: Yale University Press)

Autonomy and Humanity as an End in Itself

Carlos Gutierrez

Introduction

Kant, in the *Groundwork*, states the "humanity formulation" as follows: (FH) "So act that you use humanity, in your own person as well as in the person of any other, always at the same time as an end, never merely as a means" (Kant 4:429).[1] This principle demands that one should treat persons as ends in themselves, and never as a means. It expresses what Kant calls a Categorical Imperative, a command that one must conform to regardless of one's inclinations. There is general agreement that all minimally rational beings must be treated as ends in themselves because they have an absolute value. This absolute value is called "dignity." However, the debate focuses on what exactly gives a rational being dignity or, to put it differently, what the word "humanity" means. One prominent view, argued by Christine Korsgaard and Allen Wood, is that a rational being attains this dignity and becomes an end in itself in virtue of having the capacity to set ends. Although there is textual evidence for this view, I will argue against it in favor of an alternate view. The view that I propose is that a rational being is an end in itself in virtue of autonomy and this is based on more accurate textual evidence that complements Kant's overall moral philosophy.

But, in order to argue for the above claim, I will first explain, in section I, what kind of "end" is needed in order to ground the Categorical Imperative. In section II, I will give an account of "the capacity to set ends" reading and show why it is problematic. In section III, I will present my alternate reading that is called the "autonomy" interpretation. But in order to do this, I will first explain what Kant means by morality, freedom and autonomy. I

will then argue for the autonomy reading based on strong textual evidence and show why it is a more accurate interpretation of Kant's moral philosophy. Lastly, in section IV, I will complete my interpretation by connecting it to Kant's conception of dignity and address possible objections one might have to the autonomy interpretation.

Section I — Objective Ends vs. Subjective Ends

I want to start this discussion by giving an account of the distinction between objective ends and subjective ends. This distinction is crucial in understanding what Kant means by "humanity." In the *Groundwork* a rational will can be motivated to obey the Categorical Imperative only by a special kind of "end" (*Zweck*). Before Kant introduces the FH, he writes: "What serves the will as the objective ground of its self-determination is the end" (Kant 4:427). In other words, ends serve to determine the will—rational beings are able to set their own ends rather than having them provided by nature. In order to understand what he means in this important passage, it is crucial to explain what "ends" are.

According to Kant, a will can engage in two kinds of ends. First, a subjective end which is based on some empirical desire for an object and is called an "incentive" (*Triebfeder*), and second, an objective end, which is given by reason alone and is valid for all rational beings and is called a "motive" (*Bewegungsgrund*) (Kant 4:427-428). Further, there is a distinction between two forms of "objective." Kant does not directly make this distinction but it is implicit in his account of objective ends. They can be called objective$_1$ and objective$_2$.[2] The former refers to all ends that are grounded on rational deliberation and for this reason they are universal and objectively valid. It is important to note that objective$_1$ ends can also be subjective. The latter refers to ends that are grounded by pure reason and are valid for all rational beings, irrespective of their inclinations. In this way, this kind of end acquires an absolute value rather than a relative value because it is inde-

pendent of an agent's desires. Kant is unclear about the role of absolute value in the FH so I will discuss this further in section IV. The important point here is that humanity must be grounded on an objective$_2$ end since the Categorical Imperative must be valid for all rational beings.

Also there is a distinction between "ends to be effected" and "independently existing ends" (Kant 4:437). An end to be effected is a state of affairs that is brought into existence by a person's causality. An independently existing end is a state of affairs that already exists and constitutes a limit. The distinction here is important because humanity, i.e. the Categorical Imperative, must be grounded on the latter. However, it is important to note that not every independently existing end has the capacity to ground the Categorical Imperative. For example, when people salute their country's flag they act for the sake of an end, that is, they show a deep respect by acknowledging the symbolic value that is attached to the flag.[3] Nonetheless, even though respected objects like flags can be independently existing ends, they cannot ground the Categorical Imperative. This is because they are not objective$_2$ ends. In other words, they are not valuable for the same reasons that hold for all rational beings.

In short, if any end is to ground the Categorical Imperative it must be (1) an objective$_2$ end and (2) an independently existing end. The meaning of the word humanity must meet these two requirements and must hold for all rational beings. Once humanity meets these requirements, it must never be treated as a mere means, that is, it must be treated as an end in itself. Once we understand what the word "humanity" means in the FH, then we can understand why it has absolute value.

Section II — The Standard Reading

Many commentators have provided interpretations for the meaning of the word humanity. However, this paper will only focus on one interpretation, namely, the "capacity to set ends." This interpretation is the most prominent one.

Korsgaard argues for the "capacity to set ends" interpretation. In this interpretation only a rational agent with the mere capacity to set an end has humanity. To say that a rational being has the power to set an end is identical to the Kantian term *Willkür*: the power to choose. This is a power that all minimally rational beings have. This entails that all minimally rational beings fall under the FH. Her main source of evidence is that Kant explicitly writes in the *Groundwork*: "Rational nature is distinguished from others in that it proposes an end to itself" (Kant 4:437). Similarly, Kant writes in the *Metaphysics of Morals*: "The capacity to set oneself an end—any end whatsoever—is what characterizes humanity (as distinguished from animality)" (MM 392). This quote suggests, Korsgaard argues, that what distinguishes humans from animals is that a human can set an end by using practical reason, and this is why human action is governed by practical reason—whereas an animal's action is guided by mere instinct. Now, since human action is governed by practical reason, one can choose what maxims to act on, and given that every maxim has an end, it follows that an end is a free choice.

There are two problems with this view that I would like to address. First, it is unclear, in reading Korsgaard, to what extent morality plays a role in a rational being's capacity to set ends. Korsgaard is aware of this difficulty and addresses this point by saying that this capacity is "only completed and perfected" insofar as a rational being "responds to moral incentives" (Korsgaard 1996, p. 114). One can read Korsgaard's view as incorporating morality as a necessary condition in her capacity to set ends reading. That is to say, a capacity to set ends only has value because it inevitably leads one to morality. But in order to do this, one needs what Kant calls *Wille*: the power to self-legislate moral laws, in addition to *Willkür*. One can also interpret Korsgaard's view as only containing *Willkür*. That is, only the capacity to set ends is what makes a rational being an end in itself. This lack of firm grounding makes Korsgaard's interpretation vague. If one interprets her view as containing morality, then she has not given any direct textual evidence for this support. But, if one grants her

the latter view, the one that does not contain morality, then her view runs into the same problem that Wood's view runs into. I will discuss this problem below.

A more prominent problem with Korsgaard's view is the context of the passage in Kant's *Metaphysics of Morals* that she uses for support of her view. In the quote, Kant is primarily concerned with distinguishing humanity from animality. He is concerned with the basic capacity that distinguishes rational beings from mere instinctual drives that animals have. This is why he says that the capacity to set an end is the fundamental characteristic of human beings. Human beings have a duty to set themselves apart from animality. In other words, being different from animals is not the same as having absolute worth which is what the HF entails. Kant argues later in the *Metaphysics of Morals* that human beings have only extrinsic value in their capacity to set ends, it is only when they attain a dignity that they become ends in themselves (MM 435-6). This shows that the passage cited by Korsgaard is an inadequate characterization of the humanity that Kant is thinking about when he states the FH. For this reason, the passage cited by Korsgaard is misleading if one is trying to understand humanity in the context of the FH.

Similarly, Wood argues that humanity is a rational agent's capacity to set ends (i.e. *Willkür*). However, he goes a step further by adding other features that rational agents have and, also, argues that humanity contains all our rational capacities. However, he excludes any kind of morality. Unlike Korsgaard, Wood is unambiguous in the role that morality plays in the capacity to set ends. He clearly excludes any moral aspects from the word humanity in the FH. Wood starts off by referring to a passage in Kant's *Religion* text where he distinguishes three original predispositions: "animality," "humanity," and "personality" (R 6:26). Animality is responsible for the instinctual drives that serve for self-preservation. The personality predisposition is responsible for "the rational capacity to respect the moral law and to act having duty or the moral law as a sole sufficient motive of the will." The humanity predisposition "encompasses all our rational capacities having no

specific reference to morality" (Wood 1999, p. 118). Wood also cites Kant's *Anthropology* text where Kant subdivides humanity into the "technical predisposition" and the "pragmatic predisposition." The former allows us to deliberate what the best means are to our contingent ends. The latter "enables us not only to set ends but to compare the ends we set and organize them into a system" (Wood 1999, p. 119). It is in this "pragmatic predisposition" where the concept of rational self-love arises. This concept of rational self-love enables one to attain a self-worth and it is measured by comparison of the self-worth of others. It is this comparison that enables one to feel happy or unhappy depending on where one's self-worth lies in comparison to others, that is, whether one's self-worth is inferior or superior to others. Wood concludes that the capacity to set ends through deliberation is what actualizes humanity—and humanity includes the technical and pragmatic predispositions while excluding morality.

There are two important problems with Wood's interpretation. First, since he is using the passage in Kant's *Religion* as evidence for his interpretation, he is forced to consider why Kant grounds rational beings as ends in themselves based on humanity rather than personality. In the passage he cites, it can be argued that personality is a better candidate since morality is a necessary condition for personality. On the other hand, self-love, as oppose to morality, falls under humanity in Wood's reading. But according to Kant, self-love is the source of "diabolical vices" like envy and joy in other's misfortunes (R 6:27). For this reason, given the context in which Wood cites the evidence for his view, it seems like personality is a better candidate for a rational being to have the absolute value of dignity and be an end in itself.

The second problem is related to the first, that is, Wood's view does not account for the fact that Kant explicitly says that it is morality that makes an agent an end in itself. For example, in the *Groundwork*, Kant writes, "Morality is the condition under which alone a rational being can be an end in itself" (G 4:436). Since Wood explicitly excludes any kind of morality as a necessary condition for a rational being to be an end in itself,

this passage clearly contradicts what Kant seems to suggest in the *Groundwork*. This is a problem for Korsgaard's interpretation as well if one interprets her view as excluding morality.

The important lesson here is that one must account for morality if one is to give an account of what humanity is and why it is an end in itself. Kant's moral texts suggest that at the very least the capacity for morality is needed for humanity to be an end in itself. Thus, any interpretation that does not account for morality will be inadequate.

Section III — The Alternative Reading

In the last section I argued that the textual evidence for the "capacity to set ends" reading is not strong because it is unable to account for morality. In this section I will argue for an alternate reading, namely, the "autonomy" reading. In this reading what makes a rational being an end in itself is autonomy. I will argue, based on textual evidence, that this interpretation is correct because freedom and morality are contained in the concept of autonomy. To begin, I will first explain what Kant means by morality and freedom.

Kant defines morality in four different ways:[4] (1) right actions, (actions like not lying and helping others), (2) right actions regulated by the Categorical Imperative, (3) self-legislation of moral principles, (4) the categorical imperative itself. In the autonomy interpretation I suggest that morality (3) is contained in the concept of autonomy, I will call this morality_a. Similarly, Kant defines freedom in various senses. One sense of freedom can denote unconstrained action. Another sense is called "negative freedom" and it means actions not dictated by pre-existing conditions. Lastly, there is "positive freedom," which is deeply self-caused (G 4:447). The freedom that I want to use in my argument is positive freedom. I will call this freedom_a. In this kind of freedom a rational being is free not only because she can choose, but also because she chooses to act on reasons that that take the form of moral principles that are produced by her own will. In this

way, it is evident that morality$_a$ and freedom$_a$ are equivalent.

Now, morality$_a$ and freedom$_a$ are both contained in Kant's definition of autonomy. In other words, I argue that all three terms are equivalent. Kant defines autonomy as the will self-legislating its own moral laws by pure practical reason (G 4:440). This self-legislation of moral laws through reason originates in the agent's self and is chosen by the agent. That is to say, what makes a rational being autonomous is not only her ability to choose but that through her deliberation she chooses the right moral principles that originate from her own will. Given this definition of autonomy, both morality$_a$ and freedom$_a$ are necessary conditions for autonomy. This is why Kant writes: "Freedom and the will's own lawgiving are both autonomy and hence reciprocal concepts" (G 4:450). "Freedom" in this passage is consistent with the meaning of freedom$_a$ and the "will's own lawgiving" is essentially morality$_a$. This passage suggests that both concepts are equal to each other and, in turn, both are equivalent to Kant's conception of autonomy.

So far I have shown how morality$_a$ and freedom$_a$ and autonomy are all equivalent to one another. Now I will argue that a rational being is an end in itself in virtue of having autonomy. Consider when Kant writes:

(M) "Morality is the condition under which alone a rational being can be an end in itself" (G 4:435).

(F) "Freedom, only freedom alone, makes it that we are ends in ourselves" (Guyer 2000, p. 156).

It might seem like Kant is contradicting himself here since in (M) only morality makes a rational being an end itself whereas in (F) only freedom makes a rational being an end in itself. However, since I have shown above that morality$_a$ is the same as freedom$_a$ and that both are contained in the concept of autonomy, it follows that (M) and (F) are essentially saying the same thing. In other words, in my reading of Kant, (M) and (F) show that it is autonomy that makes a rational being an end in itself. Again, the contexts of

these quotes were carefully chosen and are consistent with the type of morality and freedom that I have been writing about.

Further, Kant also states that autonomy is a sufficient condition for a rational being to be an end in itself. He writes in *Naturrecht Feyerbend*: "Freedom is not only the highest but also the sufficient condition" for a person to be an end in itself (Guyer 2000, p. 157). This passage shows that only autonomy is needed for a rational being to be an end in itself. Now, since the FH is the Categorical Imperative, it follows that anyone that has autonomy must be treated as an end in itself. A rational being that is free cannot use her freedom in order to treat other free beings as a mere means. If she does treat other free beings as a mere means then she will be destroying their freedom that is intrinsically valuable.

In light of this one can see how the other two Categorical Imperatives are connected to the FH and help preserve the humanity (i.e. autonomy) in rational beings. Kant says that the three Categorical Imperatives are one in the same. The Formula of Universal Law (FUL) requires one to act only on maxims that can be willed universally (G 4:421). Since autonomy is intrinsically valuable, this formulation guarantees that the autonomy of other beings gets preserved. Also, consider the third Categorical Imperative, the Kingdom of Ends formulation that is "the principle of every human will as a will universally legislating through all its maxims" (G 4:432). This formulation also helps preserve the intrinsic value of autonomy. Every autonomous rational being in the Kingdom of Ends is brought together by common laws that have been self-legislated. The main point here is that all three formulations of the Categorical Imperative are grounded on the concept of autonomy and help preserve it. This in turn, I think, strengthens my interpretation. Kant will go on to argue that the Categorical Imperative is made possible by autonomy in section three of the *Groundwork*. Given the crucial role that autonomy plays throughout Kant's project, the autonomy reading is compelling. Nonetheless, the issue of value needs to be further explained. In the next section I will explain how my interpretation addresses the concept of value.

Section IV — Autonomy and Dignity

In the last section, I explained how the autonomy interpretation was correct based on textual evidence. In this section I want to show how my interpretation accounts for the concept of value. Also, I want to address two important objections one might pose against my interpretation.

Kant seems to suggest that dignity is an essential condition for something to be an end in itself, even though he does not directly argue for this claim. Kant writes before introducing the FH:

> But suppose there were something the existence of which in itself has an absolute worth, something which as an end in itself could be a ground of determinate laws; then in it, and in it alone, would lie the ground of a possible categorical imperative, that is, of a practical law (G 4:428).

The crucial point in this passage is to figure out what it means for a person to have "absolute worth" in virtue of being an end in itself. Kant seems to suggest that an end in itself has to have a special value that cannot be furthered or brought about by any action. He later claims that dignity is this special value and it is a necessary condition for something to be an end in itself (G 4:435, MM 434-5). He contrasts dignity's "absolute worth" with price's "relative worth" (G 4:435). Something that has relative worth is a subjective end. A subjective end has relative worth because it depends on a particular inclination that a rational being might have. For example, a cup of coffee has value for me only because I desire it. In other words, the cup of coffee has value only insofar as someone has a desire (i.e. inclination) for it. Its value is dependent on my desire for it and, for this reason, has relative worth. Conversely, an end in itself has "inner value" and this value is an absolute value because it is not dependent on a rational being's inclinations. This makes an end in itself an objective$_2$ end (see section 1 for definition). If something is an objective$_2$ end and has absolute value, then it has dignity.

Given the above reasoning, if autonomy is the condition that makes humanity an end in itself, then it must have an absolute value (i.e. dignity). In other words, it must be an objective$_2$ end and an independently existing end. If one interprets Kant in this way, then dignity must be a necessary condition for autonomy since it is an end in itself and must be a "necessary end for everyone" (G 4:428). It is shown that autonomy has dignity when Kant explicitly writes: "Autonomy is thus the ground of the dignity of a human and of every rational nature" (G 4:436). The word "autonomy" in quotes denotes the same definition that I've been using because in the same paragraph he describes a free rational being that self-legislates her own universal moral laws in the Kingdom of Ends. This passage shows that autonomy is the ground for the Categorical Imperative because it emerges from a rational being in virtue of her freedom. As I argued in the last section, all three Categorical Imperatives are design to protect and preserve the autonomy of a person. Now one can see that it needs to be preserved because it is the source of dignity.

Furthermore, consider when Kant writes in the *Naturrecht Feyerabend*: "Man namely is an end in himself, he can therefore have only an inner value, i.e. dignity, on which no equivalence can be set…. The inner value of man rests on his freedom, that he has his own will" (Guyer 2000, p. 152). This passage further supports that dignity rests on autonomy. The word "freedom" is meant to be freedom in the positive sense (freedoma), therefore it is equivalent to the word autonomy. Kant places autonomy above anything else "on which no equivalence can be set." Since nothing else is more valuable, it makes sense that autonomy should be the necessary and sufficient condition for a rational being to have dignity and be an end in itself. The two quotes above provide compelling textual evidence that autonomy has dignity.

Now that I have argued for the autonomy interpretation, it is important to consider two objections. One objection is that I have not taken a firm stance on the role that morality takes in the autonomy interpretation. In other words, under my interpretation, does morality have to be actualized by rational beings or is

it sufficient to say that they have a mere capacity for morality? In the *Groundwork* Kant writes: "morality and humanity, in so far as it is capable of morality, is that which alone has dignity" (G 4:435). This passage suggests that it is only the mere capacity for morality that gives a person dignity.[5] However, Kant begins the *Groundwork* by stating that a good will is the only thing good without qualification (G 4:393). A good will, according to Richard Dean, is "the will of a being who is committed to acting morally" (Dean 2006, p. 24). In other words, a good will is not only the capacity for morality but also the actualization of morality. Moreover, Kant seems to suggest that humanity and a good will are equivalent.[6] The overall problem is now clear: is the capacity for morality sufficient for a rational being to be an end in itself or does she need to act on this capacity? This problem is compelling and the autonomy interpretation needs to account for it.

I want to respond to this problem by arguing that one does need to act on morality in order to attain dignity. Consider a quote that I cited earlier in Kant's *Naturrecht Feyerabend*: "Freedom is not only the highest but also the sufficient condition" for a person to be an end in itself (Guyer 2000, p. 157). Kant is essentially saying that autonomy is a sufficient condition for a rational being to have dignity. But it must be remembered that autonomy contains the moral concept of a will self-legislating its own laws. Now, consider this passage in the *Groundwork*: "the dignity of humanity consists in just this capability, to be universally legislative, if with the proviso of also being itself subject to precisely this legislation" (G 4:440). The keyword here is "capability." Capability is ambiguous. For example, in theory, an amateur swimmer has the capability of winning a gold medal in the Olympics. However, a young professional swimmer who has won a gold medal in the previous Olympics also has this capability. The point I want to suggest is that when Kant uses the word "capability," he means it in the sense that it will be actualized like in the case of the professional swimmer.[7] In this way, the autonomy interpretation accounts for the incomparable worth that Kant places on the good will. However, my interpretation is different in that dignity is not

grounded on the good will alone but in autonomy.

Another objection to the autonomy interpretation is the foundational role that absolute value (i.e. dignity) has in the reading. It can be argued that one does not need to ground humanity on dignity in order to treat rational beings as ends in themselves. One treats them as ends in themselves because the Categorical Imperative commands it. For this reason, it is pointless to try to argue that dignity is needed in order to treat a person as an end in itself. Moreover, this objection is further strengthened by the fact that Kant does not directly argue that humanity needs to have dignity in order to be an end in itself. Yet, according this criticism, it should not matter because the source for this is the Categorical Imperative itself, thus dignity is not needed.

This second criticism seems to be missing the point. That is, of course the Categorical Imperative requires one to treat someone as an end in itself. But the issue is: under what condition does a rational being become an end in itself. I have argued that this condition is autonomy. This line of reasoning, in response the second criticism, leads me to say that the Categorical Imperative would not be possible unless rational beings had autonomy. The point here is that the Categorical Imperative is grounded on autonomy. For this reason, autonomy is needed, and, moreover, it has absolute value. In other words, if autonomy did not have absolute value then there would be no grounding for the Categorical Imperative. Therefore, dignity is needed.

CONCLUSION

In this paper I have argued that, on a close reading of Kant's texts, the "capacity to set ends" view is problematic. The textual evidence for it seems to fall apart once one analyzes the context in which they are quoted. The main problem with this interpretation is that it does not account for morality. This view is also inconsistent when Kant claims that morality and freedom are the sole reasons a rational being attains dignity. My reading, in contrast, takes into account the context of the quotes that were cited. In the

textual evidence that was used, I made sure to take into account that Kant uses the same words to denote different meanings. I was able to account for the terms of morality, freedom, and autonomy that play a vital role thought Kant's moral philosophy. Moreover, I also showed that autonomy is a sufficient condition for a rational being to have dignity and thus be an end in itself.

Notes

1. Citations to Kant's works will be given by volume and page number in the Royal Prussian Academy edition, which are found on the margins of the English translations that I use.
2. I owe this distinction to Allison (Allison, 209).
3. I owe this example to Wood (Wood, 116).
4. I thank Richard Dean for making these distinctions clear in a discussion I had with him in 2015.
5. For a full account of this interpretation see (Allison, 215-218).
6. For a full account of this interpretation see (Dean, 37-49).
7. Richard Dean pointed this idea out in a discussion I had with him in 2015.

Bibliography

Allison, Henry. (2011) *Kant's Groundwork for the Metaphysics of Morals* (New York: Oxford)

Dean, Richard. (2006) *The Value of Humanity in Kant's Moral Theory* (New York: Oxford)

Guyer, Paul. (2000) *Kant on Freedom, Law, and Happiness* (New York: Cambridge University Press)

Kant, Immanuel:

(A) *Anthropology from a Pragmatic Point of View*, trans. Robert B. Louden, (New York: Cambridge University Press, 2006), Ak 25.

(G) *Groundwork of the Metaphysics of Morals*, trans. Mary Gregor (New York: Cambridge University Press, 2011), Ak 4.

(MM) "The Metaphysics of Morals," in *Practical Philosophy*, trans. Mary Gregor (New York: Cambridge University Press, 2006), Ak 6.

(R) *Religion within the Boundaries of Mere Reason*, trans. Allen Wood (New York: Cambridge University Press, 1998), Ak 6.

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

Wood, Allen W. (1999) *Kant's Ethical Thought* (New York: Cambridge University Press).

A Case For Pragmatic Foundationalism

Stephen Sigl

Empirical foundationalism is a very broad and controversial topic that has engendered strategies aimed at upholding some semblance of it, as well as strategies that stand in direct opposition to it. The long-standing opponent of foundationalism is coherentism. However, in this paper I will argue that pragmatism, which is often regarded as essentially anti-foundationalist, can accommodate the metaphilosophical demand for a foundationalist approach to philosophy.

The first section of the paper will be a brief outline of the foundationalist position and some objections that have been brought against it. The two primary problems I will address are what I call the "coarse grain issue", 'how can we have confidence in beliefs generated from our senses to the extent that they might be sufficiently justifiable?', and the "fine grain issue", 'how can a belief, which is propositional, be derived from a sensation'? In this section I will also discuss Lawrence Cahoone's metaphilosophical reason for linking foundationalism to philosophy proper. In the second section I will briefly examine the pragmatic approach to philosophy and how it has come to be seen as an 'antiphilosophy'. Finally, in the last section I will argue that Robert Brandom offers a tenable pragmatic response to the concrete and metaphilosophical issues raised by the foundationalism debate, and in doing so gives us reason to reconsider Cahoone's claim that pragmatism is an 'antiphilosophy'.

Section One

The foundationalism I will be discussing relates specifically to the empirical-epistemological considerations related to Cartesian

first principles. For Descartes, these first principles are established through a deliberate mediation of contrasting inclinations towards doubt and certainty. He thought this approach to knowledge would offer a foundation for philosophical inquiry that mirrored a mathematical system like Euclidean geometry, wherein definitions, axioms, and postulates form the superstructure for further propositions (Newman 2014).

Descartes, who inherited the Platonic view that knowledge is justified true belief (JTB), sought to map this foundational structure onto our conception of empirically acquired knowledge. Modern foundationalist theories essentially seek the same goal and are characterized by a one-directional flow from sensory experience to justified belief, e.g., I touch a hot stove and subsequently develop the propositional belief that the stove is hot. According to Laurence BonJour, the foundationalist position requires that this sensory experience generate what is called an "unjustified justifier": it is sufficiently self-evident, and further beliefs can be inferred from this unjustified justifier. BonJour refers to such an "unjustified justifier" as having "a justification that does not derive from other empirical beliefs in a way which would require those beliefs to be antecedently justified" (BonJour 1985, p. 26). That is, they have some degree of noninferential justification.

BonJour's critique of foundationalism, for the sake of this essay, amounts to two issues that I have designated with the terms "coarse grain" and "fine grain". The coarse grain issue relates to the problem Descartes saw from the start of his investigations: how do we trust our sensory input to the extent that it can warrant further inferential beliefs? Perhaps I hallucinated that I touched the stove and am therefore unjustified in believing that it is hot, or even that there is a stove?

The proper discernment of (potentially) arbitrarily generated empirical beliefs from genuinely "unjustified justifiers" leads to a question of criteria. If we say that the basic units of justification that have the special status of being unjustified justifiers possess a specific characteristic or feature, we next have to ask whether or not having the specific characteristic constitutes "a cogent reason

for thinking that it is likely to be true" (BonJour 1985, p. 23). An insufficient answer to this query leads to the conclusion that further empirical beliefs are of an arbitrary nature. This opens foundationalism up to the skeptical questions of where or how a basic justifier can be *found*, let alone substantiate the realm of empirical belief. If any beliefs can be justified, there must be some existing, independent beliefs, which do not owe their justification to anything else. If we required these basic, or terminal, beliefs to be justified in the same manner in which the beliefs that are inferred from them are, then the chain of justification would be interminable and hence an infinite regress would be generated. It is not enough that these terminal beliefs be self-evident, they need also to not owe their justification to anything except sensory experience (BonJour 1985, p. 18).

The fine grain issue is more subtle and has to do with the question of how an experience, which is itself not a belief, justifies a belief. The two processes seem qualitatively incommensurate, insofar as beliefs are expressed as propositional attitudes that do not entail a logical relationship to sensation, and demand an explanation of what their causal interaction would entail; an explanation which seems to fall outside the bounds of reasonable philosophical inquiry (Davidson 2001, p. 143).

BonJour identifies three types of foundationalism: *strong*, *moderate* and *weak* foundationalisms; the focus of his critique is directed at moderate foundationalism, though he states that his argument is implicitly applicable to all three:

> For although this has often been overlooked, the very idea of an epistemically basic empirical belief is more than a little paradoxical. On what basis is such a belief supposed to be justified, once any appeal to further empirical premises is ruled out? […] How can a contingent, empirical belief impart epistemic "motion" to other empirical beliefs unless it is itself in "motion"? (BonJour 1985, p. 30)

For BonJour there are two strategies by which the foundationalist might approach the problems associated with foun-

dationalism. They can be roughly generalized as *externalist* and *internalist* strategies. For the externalist, the problem with foundationalism lies in the *truth* portion of JTB; this in turn gives rise to an emphasizing of the *belief* portion. For the externalist, it is not necessary that anyone know or justifiably believe they are in cognitive possession of the reason why a basic empirical belief is a basic justifier; the justification is essentially external to the subject's awareness. Externalism usually relies upon the reliabilist hypothesis that one is entitled to the knowledge claim of *P* if and only if *P* is true, one believes *P* is true, and said belief has been arrived at through a reliable process.

The other option is a weak foundationalism (or internalism) that co-opts the most useful aspects of coherence theory in order to circumvent the problem of justification. Coherentism is a strategy that situates justification within a network of mutually sustaining beliefs, and is itself free of the infinite regress of justification that plagues foundationalism. This form of coherentism is a holistic and nonlinear conception of justification for which BonJour lists three objections. The first is that coherentism allows for arbitrary pluralism, i.e., allows there to be multiple systems of belief within the same subject, the choice between which will be arbitrary. The second is that empirical knowledge is deprived of any input from the non-conceptual world, hence limiting the subject's ability to describe that world. Finally, BonJour claims that such a theory will be seemingly unable to "establish an appropriate connection between justification and truth unless it reinterprets truth as long-run coherence" (BonJour 1985, p. 25).

This all leads up to an important question: if foundationalism is wrought with so many difficulties, why would we want to maintain it? According to Lawrence Cahoone, foundationalism is central to philosophy because it entails *realism*, the belief in an objectively existing external world, and philosophy would not be philosophy if it abandoned realism:

> I will suggest that foundationalism of a kind cannot be abandoned on pain of self-inconsistency or of ceasing to

> do philosophy at all. Attempts to limit philosophy, however sophisticated, violate philosophy's nature. If philosophy is primarily ultimate inquiry, then philosophy cannot accept limited validation as sufficient, nor can it accept the validation of judgments in terms of some norm other than truth (e.g., practical necessity or goodness). (Cahoone 1995, p. 11)

This is essentially a metaphilosophical argument because it is talking about philosophy in general and how it ought to operate. For Cahoone, foundationalism is the issue that makes apparent philosophy's necessary commitment to realism. Cahoone utilizes a few novel strategies in his defense of foundationalism. To begin with, the classification of philosophy as "ultimate inquiry" indirectly addresses the infinite regress problem; infinite regress is problematic, but not necessarily vicious according to Cahoone, and understood properly, it is actually an unavoidable consequence of doing philosophy. In light of this, Cahoone admits that philosophy opens itself up to ultimate inquiry as well as the corresponding demand for ultimate validation and that, in this respect, it cannot issue judgments that it might satisfactorily deem assertively valid—philosophy, by its very nature falls outside the scope of objectivity necessary for such an assessment. Cahoone offers two reasons why this paradoxical situation warrants an endorsement of foundationalism. First, he insists that although it might seem necessary to place limits on inquiry, it is the task of philosophy to question these proposed limits. The fact that final validation does not exist is not problematic; it is merely a necessary component of the concept of ultimate inquiry (Cahoone 1995, p. 326). Secondly, Cahoone insists that his thesis is established negatively, i.e., by process of elimination: nonfoundational attempts to resolve the foundationalist problem are inadequate on his view and antiphilosophical (or pragmatic) attempts to dissolve it fail because they remain within the philosophical discourse and are therefore subject to self-reflective inconsistencies. There is, in effect, no principled escape from "the burdens of philosophical

validation that can legitimately regard itself as achieving a validity philosophy ought to recognize" (Cahoone 1995, p. 11).

The negatively established thesis is a provocative strategy that causes us to question the absoluteness of BonJour's *strong*, *moderate* and *weak* classification—it essentially evades the issues of the argument by pointing to the bigger picture of what is lost if foundationalism is relinquished: a legitimizing ground for further inquiry. It is metaphilosophical not because it is talking about how philosophy has, or is operating, but how it ought to operate. Metaphilosophy is normative, then, and therefore requires a discussion of normativity. But before that we will have to look at the pragmatist response to the problems raised by foundationalism and see why Cahoone takes such a strong stand against pragmatism.

SECTION TWO

For this section I will discuss pragmatism not only as an ethos that emphasizes practice over theory (as evinced by the pragmatic motto: "knowing *that* is a matter of knowing *how*"[1]), but also as an ethos that emphasizes the social context of knowledge. In the first part I will discuss Peirce's rejection of Cartesian skepticism and in the second part I will refer to Rorty's defense of the communal nature of knowledge. It is important to keep in mind that even though Rorty drew more heavily on James and Dewey than he did on Peirce, his pragmatism would certainly endorse Peirce's injunction that:

> We individually cannot reasonably hope to attain the ultimate philosophy which we pursue; we can only seek it, therefore, for the community of philosophers. Hence, if disciplined and candid minds carefully examine a theory and refuse to accept it, this ought to create doubts in the mind of the author of the theory himself. (Peirce 1868, p. 71)

My point is simply this: there are divergent forms of pragma-

tism. It is not a homogeneous ideology, although there are certain common threads that allow us to speak of a modern pragmatism while respecting the fact that its partisans are not in agreement on every issue.

To understand the epistemological position taken by modern-day pragmatists like Richard Rorty it is useful to compare and contrast it with the previously mentioned coherentist model. I will not attempt to give a thorough examination of coherentism; instead, I will merely refer to Donald Davidson's proposal as propounded in *Subjective, Intersubjective, Objective* insofar as it is a formal coherentist theory which is motivated by a desire to preserve realism.

Davidson puts this desire at the forefront of his argument by stipulating that his program relates expressly to 'knowledge' and not truth: 'Truth' itself is not to be defined in terms of coherence and belief—truth itself is transparent (it corresponds to "the way things are") and ought to be treated as a primitive concept (Davidson 2001, p. 139). Davidson does not characterize pragmatism as an 'antiphilosophy' in the way that Cahoone does—up to a certain point he emphasizes a connection between his view and that of Richard Rorty's:

> What distinguishes a coherence theory is simply the claim that nothing can count as a reason for holding a belief except another belief. Its partisan rejects as unintelligible the request for a ground or source of justification of another ilk. As Rorty has put it, 'nothing counts as justification unless by reference to what we already accept, and there is no way to get outside our beliefs and our language so as to find some test other than coherence'. (Davidson 2001, p. 141)

The only divergence Davidson sees between coherentism and Rorty's pragmatism is that Rorty is not interested in seeking a means by which we may speak of an objective world that is not of our own creation (Davidson 2001, p. 141).

Davidson's resistance should be read as a tacit unwillingness to sign off on the possibility of any anti-realism that might be derived from Peirce's rejection of Cartesian skepticism. According to Peirce, Descartes' universal doubt founders for a variety of reasons: the foremost one being that any attempt at total skepticism is bound to take for granted some propositions, either by way of overlooking them or because they are necessary to the process of doubting the specific proposition designated for skeptical inquiry. The Cartesian method ignores the contextual nature in which questionable propositions arise, "We cannot begin with complete doubt. We must begin with all the prejudices which we actually have when we enter upon the study of philosophy. These prejudices are not to be dispelled by a maxim, for they are things which it does not occur to us can be questioned" (Peirce 1868, p. 70). This argument takes on a key significance for Rorty because Descartes' project (as a solitary exercise) does not take into account the social/linguistic conditioning that has engendered (or tacitly substantiates) the proposition(s) being questioned, let alone addresses the tenability of a presumed 'solitary' forum in which they are to be questioned. According to Cahoone, Rorty uses this argument to offer an account of knowledge that is anti-realist insofar as it makes the validity of our judgments relative to vocabulary, context, community and culture:

> For us [pragmatists], all objects are always already contextualized. They all come with contexts attached… So there is no question of taking an object out of its old context and examining it, all by itself, to see what new context might suit it… Once one drops the traditional opposition between context and thing contextualized, there is no way to divide things up into those which are what they are independent of context and those which are context-dependent… there are no candidates for self-subsistent, independent entities save individual beliefs… But these are very bad candidates indeed. For a belief is what it is only by virtue of its position in a web [of belief]. (Rorty 1991, pp. 97-98)

According to Rorty, philosophy since Plato has been dominated by the view that the mind is a 'mirror of nature' and the conviction that philosophy is the discipline that grounds other disciplines because it investigates the ways in which thought and language represent the mind-independent world. For Rorty, the idea of privileged representations ought to be abandoned in favor of the view that all that exists to ground inquiry is social discourse. Hence, concepts like "truth" and "justification" are merely semantic components of conversational practices.

For Cahoone, Rorty's attempt at dissolving the questions of epistemology to social conventions is problematic because he is using the language and method of philosophy in order to make his case; i.e., Rorty is having his cake and eating it too. Cahoone insists that this is untenable: there is no such thing as a principled abstention from philosophy because said abstention, in one way or another, is always seeking some sort of philosophical validation (this is not the case for unprincipled or indifferent forms of abstention). For Cahoone, Rorty's ingenuity and attention to semantics is not only an attempt to dispose of philosophy: it is a final fleshing out of pragmatism's inherent anti-foundationalism. Because of this, Cahoone designates pragmatism as an 'antiphilosophy'.

Whether this is a deserved appellation is debatable: Peirce was in fact a realist to the extent that he assumed that the actual world was something that could eventually be apprehended progressively by way of the scientific method. What we do get from pragmatism, however, is a new way to approach the 'coarse grain issue' of justifying the validity of empirical belief: we can take the pragmatist's rejection of Cartesian skepticism at face-value, and in doing so substantially mitigate the severity BonJour attaches to the need for 'basic justifiers'.[2] This line of reasoning does not seem to preclude the necessity of Cahoone's metaphilosophical argument. However, it still leaves the fine-grain issue of the potentially incommensurate natures of beliefs and experiences still open.

Section Three

It would not do justice to the subject to merely argue against Cahoone's assertion of pragmatism being an antiphilosophy. Instead, I will try to use the pragmatic ethos as a point of entry into the foundationalism debate.

Hopefully, two basic opposing strategies are forming in the reader's mind. The first is Cahoone's solution to the coarse grain and fine grain issues, achieved by positing the metaphilosophical necessity of foundationalism. On this view, we can try to solve these problems, but acceptance of philosophy as a discipline entails that we have to live with them. The other strategy is Rorty's dismissal of the urgency of these problems by dismissing the metaphilosophical necessity of foundationalism.

In order to understand the third option, rational pragmatism, it is important to see that there is a normativity conveyed in both arguments and to look at where this normative weight might be conveyed in a modern pragmatic setting. My view is this: by attacking the coarse grain and fine grain issues from a pragmatic point of view, Brandom is essentially arguing for the same metaphilosophical necessity of foundationalism that Cahoone endorses.

To set the stage for Brandom's relationship with the issue of foundationalism, it is important to recognize that he "resolutely oppose[s] what Wilfrid Sellars calls the 'Myth of the Given'", which is the idea that there is an unmediated sensory experience that might be construed as a basis for "knowledge" or "belief" prior to the acquisition of the learnt application of concepts (Brandom and Penco 1999). This insistence on mediation conveniently circumvents antirealism while allowing for the 'ultimacy' of philosophical inquiry advocated by Cahoone:

> Foundational knowledge, knowledge of reality that is not open to question, is inaccessible. On this I am in agreement with nonfoundationalists and antiphilosophers alike. The question is whether, absent such knowledge, philosophical knowledge can be justified through the strategy of

> abandoning that element of foundationalism that I earlier claimed was inexpugnable, namely, ultimacy. (Cahoone 1995, p. 154)

Brandom does not abandon the inferential nature of knowledge that dogs the traditional foundationalist perspective. Instead, he acknowledges that there are, in fact, conceptual contents being articulated in assertions. These conceptual contents, however they might be determined and upheld, are essentially standing in the place of the immediate and incorrigible judgments the foundationalist historically placed at the center of inquiry. For Brandom, the Myth of the Given amounts to a strongly foundationalist perspective; therefore, instead of using concepts as a means of distancing himself from the foundationalist position entirely, he views them as a component of a discursive disentanglement from the *strong* foundationalist position.

It could be argued that this is just a more sophisticated articulation of the traditional foundationalist outlook. However, this argument would assume an underlying commensurability between the three forms of foundationalism defined by BonJour.[3] Brandom's project is one that adopts pragmatism both as a socially conditioned ground by which semantic contents are mediated, *and* as an explicitly rational perspective that holds subsequent conceptual contents to normative standards.

The strength of the primitive nature of concepts is tied to rationality insofar as it is the normativity entailed by rationality that allows for the discrimination between good and bad inferences. Inference is a type of activity that both conforms to the pragmatist motto "knowing that is a kind of knowing how" and invites normative standards of justification (Brandom 1994, p. 91). For Brandom, the only way in which norms become relevant to conceptual content is through inference and it is specifically propositional contents that are conceptually articulated. Both aspects of the pragmatist motto are subject to normative standards—"knowing that" is merely an explicated version of "knowing how".[4] The conceptual universality entailed by this

claim is substantiated by Brandom's assertion that propositional contents can be (and often are) inferred by concept-bearers onto the practices of non-semantic animals (Brandom 1994, p. 83). The fact that such conceptual content might exist within the non-semantic animal's consciousness gives us reason to believe that concepts transcend the anthropocentric realm, yet their explication is wholly dependent on the rationality of human beings.

It is in this light that we can replace the fine grain issue of the incommensurate nature of experiences and beliefs with the more tenable account of commensurability between semantic content and conceptual content—but this only flies if we allow pragmatism to act not as a break from the foundationalist position but as a *means* by which a 'weak' foundationalism establishes itself in a way that is not merely an incorporation of coherentism. The coherentist position precludes the subject of any input from the non-conceptual world, but Brandom does not deny the existence of the non-conceptual world. His response, instead, is one that resonates with Cahoone's thesis of questioning the limits of inquiry, i.e., "What are the limits of conceptualization? Where does the non-conceptual world end, and we, as rational beings, begin?"

In regard to the coarse grain issue, which is coextensive with the problem of infinite regress, the pragmatic ethos then acts as a ground for the reappraisal of internalist and externalist strategies.

Brandom defines knowledge claims attendant to foundationalism as being inferentially conditional claims, where if one wishes to commit to the assertion of statement q one must be justified in asserting the antecedent p (in a conditional where q is the consequent). At this point a whole host of regressive problems present themselves: the interlocutor may not be adequately entitled to the assertion of p, or the assertion of p is plainly not correct within the context of being an antecedent to the consequent q. The questioning of the assertion of p relates specifically to the regress problem of foundationalism, "Classical foundationalism considers only justifying in the narrow sense of an inferential activity, not in the broader sense of vindication that includes the communicational dimension appealed to by deferential endorsing

(the authority of testimony)" (Brandom 1994, p. 204).

Basic knowledge claims are based on "the practices that institute the significance characteristic of assertional performances and the status characteristic of assertional commitments must be inferential practices" (Brandom 1994, p. 158). The power of inference, however, is bound to the fact that these statements arise in an interpersonal context which is indicative of social practices, and the validity of inference itself is derived within the linguistic game of giving and asking for reasons. Assertions entail commitment and entitlement to those commitments because they are acknowledged by those who attribute and acknowledge those commitments (Brandom 1994, p. xiv). Roughly speaking, we return to pragmatism not as an attack on the idea of *grounding* but as a *ground* for the defense of the *assertion* of *p*. A regress only occurs if entitlement to premises is not properly secured.

Brandom borrows from Dewey the claim that the traditional foundational regress is due to "a kind of platonism or intellectualism [in philosophy] that saw a rule or principle… that is[,] or could be[,] made conceptually or propositionally explicit, behind every bit of skillful practice" (Brandom 2012). Instead of this intellectualism, counters Brandom, we must begin with a notion of taking or treating inferences as correct (in the logical sense) within the context of practice (Brandom 1994, p. 26). Once this is done, we are obliged to not only entertain Dewey's fundamental pragmatic approach, "which emphasizes the implicit context of practices and practical abilities that forms the necessary background against which alone states and performances are intelligible as explicitly contentful believings and judgings;" we are also obliged to supplement it with Brandom's insistence that this context of practice is necessary to the game of giving and asking for reasons (Brandom 2012).

Brandom justifies this claim by stating that this precondition is what allows us to explicate the claim "if *p* then *q*" in the form of a conditional. That is: the endorsement of logical expression (a conditional) is in actuality an explicit rendering of a practical inferential practice which has an ultimately positive justificatory

status, regardless of whether it is *actually* ontologically justified. It is for this reason that Brandom states that the approach to logical semantics must answer to pragmatics. The *prima facie* nature of these entitlements acts as an endorsement of the proprietary justification of the antecedent assertion (Brandom 1994, p. 204).

For Brandom the core of discursive practice is constituted by the linguistic game of giving and asking for reasons. In this respect, discursive practice is implicitly normative: it entails an assessment of the correct or incorrect nature of moves, and this entailment is grounded in the proprieties on the part of the practitioners. The inferential practice that confers contents of this kind is comprised of not only first-person reasoning but also third-person attributions and assessments of it—and both aspects are essential to it (Brandom 1994, p. 158).

This dissolves the epistemic regress problem of foundationalism into the folds of pragmatic theory. Normativity is the twist: do we think of normativity differently to make this work? Is it a process of redefinition then? And is this process of redefinition then analogous to the process of redefinition that foundationalism itself must undergo in order to accommodate the pragmatist's interrogation of its merit? Only in the sense that we must think of normativity expressly in the Kantian sense of "acting according to a conception or a representation of a rule, rather than just according to a rule" (Brandom 1994, p. 31). Brandom stresses this point because it allows for the pragmatic distinction of a subject acting according to her understanding or grasp of the rules—the rules, in a sense, are mediated by the rationality of the subject as they are applied in practice. This distinction allows for the assessment of performances. It is essentially a type of normativity that entails both commitment and entitlement. "The idea is, then, that the evolution of a linguistic interchange or conversation can be thought of as governed by implicit norms that can be made explicit (by the theorist) in the form of a score function" (Brandom 1994, p. 181). Conversational 'moves' are deemed correct or incorrect based on the score; though the appropriateness of it is grounded on the arbitrariness of social conventions. For example, a whole

society can be incorrect in its assessment of particular objects, but the structure of entitlements and commitments nonetheless adheres to a non-arbitrary, rule-oriented structure. Additionally, this structure is dynamic to the extent that it offers the speaker enough room to proffer accounts that differ from the communal prejudices, as long as said speaker adheres to the fundamental tenets of discourse (i.e., the rules attendant to commitment and entitlement).

For Brandom, the semantic explanatory strategy that prioritizes inference as its basic concept is at odds with the strategy that takes representation as its basic concept (Brandom 1994, p. xvi). So far I have argued, alongside Brandom, that there ought to be an inferential point at which a foundational regress should stop and that this point coincides with the ethos engendered by the pragmatic rejection of Cartesian skepticism. This is precisely the aspect of pragmatism that has generated the notion that it accords with what we might call 'common-sense'. It is important to point out, however, that Rorty's denial of the mirror of nature (the representational theory of perception) is at odds with this appeal to common sense in that it rejects the commonsense notion of thought as representation. On Brandom's view, the representational dimension of thought as a component of intentionality is significant primarily because it allows for an understanding of intentional contentfulness—there are commitments that reach beyond sequences of representations (Brandom 1994, p. 69).

If we consider a situation where a subject, Vera, is cognitively aware of a piece of paper that is sitting on her lap—the normative oughts (and attitudes/dispositions toward them) apply to the tokens of the proposition as well as to the types in her thinking: "this piece of paper is sitting on my lap." There is an analogous normative activity occurring in each. Both are mediated semantically. There is no non-linguistic thought that can amount to, or even entail, "this piece of paper is sitting on my lap." This is not to argue the claim that thought is *reducible* to language, but instead the weaker claim that thought, insofar as it is intentional, *supervenes* on language. Making the leap from an experience to

an observation report is what designates an empirical knowledge report as an "unjustified justifier" that acts as a regress-stopper; this justifier is then subject to defeasibility (Brandom 1994, p. 222).

The issue of defeasibility is tied to a codependence between internalism and reliabilism—reliabilism being a legitimate outgrowth of what we have already established as the fundamental pragmatist mindset. If we take the Goldman fake barn example (in which the subject sees an actual barn and registers it as being an actual barn despite the fact that this barn is located in an area filled with barn-facades) as being a Gettier-style challenge to the JTB view of knowledge, we can still take the pragmatic tack that an inability to tell a real barn from a cunning duplicate does not in any way disqualify us from being reliable reporters of barns.

This reliability is inferred upon us by others who are bound by the same constraints, and exemplifies the Kantian normative distinction described earlier regarding the importance placed on capacity and attitude. Claims are determined as knowledge by attributors, regardless of the claimant's attitude toward the reliability process that engendered the claim (Brandom 1994, p. 219).

For Brandom this retreat to reliabilism grounds our empirical intuitions without forfeiting the endorsement of inference; in other words, the citing of reliabilism as a mechanism for identification is inferential insofar as it warrants the acknowledgment of an empirical claim that is based on conceptual content that can be verified by members of the community who share the same conceptual capacities (Brandom 1994, p. 225).

As I stated at the beginning of this essay, foundationalism is a broad and complex topic. I hope to have adequately demonstrated two things: the first is that pragmatism is not necessarily anti-realist and that it should be considered as a valid ground for interacting with the problems engendered by foundationalism. The second is that when we speak of a case for pragmatic foundationalism, we are talking about a pragmatism that acknowledges rational commitments. In this way, we can return to the realism of Peirce and the understanding that "what is *really* a good reason, as opposed to just what is treated as one—comes at the end of

the story. It is not something that can be understood a priori and imposed as a constraint at the outset" (Brandom 1994, p. 253).

Notes

1. I.e., knowing what a term or concept means involves being able to properly use it.
2. Though an argument against systematic skepticism is not unique to pragmatism.
3. This is an implication that BonJour invites.
4. Brandom claims that the process of determining the normative account of propositions based on the notion of norms existing in practice is central to pragmatism.

Bibliography

BonJour, Laurence. (1985) *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press)

Brandom, Robert. (1994) *Making it Explicit: Reasoning, Representing, and Discursive Commitment* (Cambridge, MA: Harvard University Press)

Brandom, Robert and Penco, Carlos. (1999) "Interview with Robert Brandom," *Epistemologia*, pp. 143-150.

Brandom, Robert. (2012) "From German Idealism to American Pragmatism – and Back," *Cambridge Pragmatism: A Research Workshop*. Talk.

Cahoone, Lawrence. (1995) *The Ends of Philosophy* (New York: State University Press)

Davidson, Donald. (2001) *Subjective, Intersubjective, Objective* (Oxford: Oxford University Press)

Newman, Lex. (2014) "Descartes' Epistemology", *The Stanford Encyclopedia of Philosophy* (Winter 2014 Edition), Edward N. Zalta (Ed), URL = <http://plato.stanford.edu/archives/win2014/entries/descartes-epistemology/>.

Peirce, Charles Sanders. (1868) "Some Consequences of Four Incapacities" in: Haack (Ed), *Pragmatism, Old and New: Selected Writings*, pp. 69-106 (Amherst, NY: Prometheus Books)

Rorty, Richard. (1991) *Objectivity, Relativism, and Truth* (Cambridge: Cambridge University Press)

CONTRIBUTORS

Jaewon Choe. B.A. Honors, Philosophy, University of California, Los Angeles. Ethics, Epistemology, Moral Language, conceptions of understanding, Identity Politics, Gender and Race.

John Dail. B.A. Philosophy, University of California, Los Angeles. Ethics, Social and Political Philosophy, Philosophy of Law. John intends to apply to Ph.D. programs in the Fall of 2015.

Jonathan Fry. B.A. Mathematics, B.A. Philosophy, California State University, San Bernardino. Logic, Philosophy of Mathematics, Philosophy of Science, Philosophy of Language. Jonathan will be applying to Ph.D. programs in the upcoming winter.

Ben Futernick. B.A. Philosophy, California State University, Los Angeles. Philosophy of Science, Philosophy of Physics, German Idealism, Kant, and Metaphysics. Ben will be attending law school in Fall 2015.

Rena Goldstein. B.A. Philosophy and B.A. Humanities, University of Colorado, Boulder. M.A. Education, California State University, Los Angeles. Philosophy of Language, Metaphysics, Logic, Political Philosophy, Philosophy of Education. Rena plans to apply to Ph.D. programs in the Fall of 2015.

Carlos Gutierrez. B.A. Philosophy, B.A. English, University of California, Irvine. Bioethics, Ethics of Science and Technology, Moral Theory, Philosophy of Science, Kant. Carlos plans to apply to Ph.D. programs in the Fall of 2015.

Caleb Hahne. B.F.A. Caleb was kind enough to provide the covert art, a piece entitled "Howling Winds." Caleb has been featured in various publications, including *Juxtapoz* and *Westword* Magazines as an artist to watch. calebhane.com

Matthew Hart. B.A. Philosophy, English minor, California State University, Los Angeles. History of Philosophy, Modern Philosophy, Contemporary Philosophy, Philosophy of Science. Matthew plans to apply to Ph.D. programs in the future.

Bree Lacey. B.A. Philosophy, B.A. Communication, California State University, Los Angeles. Feminist Philosophy, Queer Theory, Transgender Studies, Disability Studies, Media Studies. Bree will continue challenging systems of oppression and interrogating binaural ideological structures.

Austin Mitzel. B.A Music, Thomas Edison State College. Political Philosophy, Social Theory, Metaphysics, and the History of Philosophy. Austin intends to pursue a Ph.D. in Philosophy soon after his M.A.

Walter Orozco. B.A. Philosophy, California State Polytechnic University, Pomona. Metaphysics, Philosophy of Time, Philosophy of Mind, Philosophy of Logic, Philosophy of Science, Ethics. Walter intends to apply to Ph.D. programs.

Nefeli Ralli. B.A. Philosophy, Psychology and Education, M.A. Ethics, University of Athens. Ethics, Meta-ethics, Philosophy of Language, Social and Political Philosophy. Nefeli plans to apply to Ph.D. programs in the near future.

Stephen Sigl. B.A. Philosophy, California State University Los Angeles. Continental Philosophy.

Ashley Tarin. Currently pursuing a B.A. in Philosophy at California State University, Los Angeles. Existentialism, Phenomenology, Feminist Philosophy, Social and Political Philosophy, Philosophy and Literature. Ashley plans to apply to the M.A. Philosophy program at California State University, Los Angeles in Fall 2015.

Master of Arts in Philosophy
California State University, Los Angeles

The Department of Philosophy at California State University, Los Angeles offers a program of study leading to the Master of Arts degree in Philosophy. The program aims at the acquisition of a broad background in philosophy. It is designed for those preparing for further graduate study or community college teaching, and for self-enrichment. Although the department is analytically oriented, it encourages work in other areas, for example Asian philosophy, feminist philosophy, and the interaction between European and Anglo-American thought. The Department includes faculty members with diverse backgrounds and interests actively working in a wide range of philosophical specialties. Classes and seminars are small with a friendly, informal atmosphere that facilitates student-faculty interaction.

The academic programs in philosophy at California State University, Los Angeles are intended to engage students in philosophical inquiry. They aim to acquaint students with noteworthy contributions by philosophers to the tradition; to explore various philosophical issues, problems, and questions; to provide students with principles of inquiry and evaluation relevant to the many areas of human activity, such as science, law, religion, education, government, art, and the humanities; to develop in them skills of analysis, criticism, and synthesis needed for advanced work in various scholarly fields; to encourage the development of skills and attitudes leading to self-reflection and life-long learning.

Philosophy in Practice
Submission Information

Each of the student contributors was specially selected to submit a paper for this issue of Philosophy in Practice by one or more faculty members in the Department of Philosophy at California State University, Los Angeles. All writers are currently either students in the master's program of philosophy or undergraduate majors in philosophy. All philosophy students at California State University, Los Angeles are eligible for nomination, and those who were chosen to contribute have demonstrated a superior ability to develop and compose works of advanced philosophical writing.

For more information on *Philosophy in Practice*, please contact:
philosophyinpractice@gmail.com