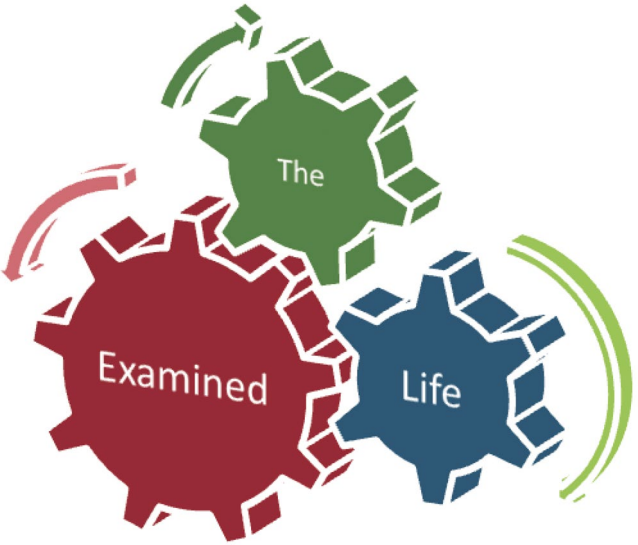
Philosophy in Practice

VOLUME 4 – SPRING 2010



CALIFORNIA STATE UNIVERSITY, LOS ANGELES
DEPARTMENT OF PHILOSOPHY

Philosophy in Practice

VOLUME 4 – SPRING 2010

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

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

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

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

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

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

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

Jennifer Faust (1993–), Ph.D. University of Illinois, Chicago. Epistemology, Philosophy of Biology, Metaphysics, Ethics of the Human Genome Project, Philosophy of Race

Ann Garry (1969–), Ph.D. University of Maryland. Feminist Philosophy, Philosophical Methodology, Epistemology, Applied Ethics, Wittgenstein, Philosophy of Law

Ricardo J. Gómez (1983–), Ph.D. Indiana University. Philosophy of Science and Technology, Philosophy of Mathematics, Kant, Latin American Philosophy

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

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–). 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

Emeritus Professors

Sidney P. Albert (1956–1979). Aesthetics, Ancient Philosophy.

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.

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

Professor Spotlight: Ann Garry

Colleagues refer to her as "an intellectual revolutionary," which causes her to blush without reason. An interactive and engaging teaching style and a cornucopia of writing, activism and academic achievements and awards commemorate a remarkable teaching career from which Dr. Ann Garry will completely retire, relinquishing the last of her teaching commitments at Cal State L.A. in 2011.

A pioneering feminist, Dr. Garry joined Cal State L.A. in 1969 to teach, but soon found herself and colleague Sharon Bishop working hard to build the philosophy department up into what it is today. Expressing a deep fondness for her colleagues and pride in their work, Dr. Garry recalls the years she and Dr. Bishop chaired the department (over 20) and how pleased she is with the recognition it has received nationwide for all that it has become.

Dr. Garry also authored an abundance of articles, reviews and scholarly presentations, and co-edited *Women, Knowledge, and Reality: Explorations in Feminist Philosophy*. She has chaired and organized panels for professional workshops, received numerous national and CSULA grants, and served on a variety of professional boards. She was the recipient of the Fourth Annual Distinguished Women's Award in 2006, and spent 2007 exploring the landscape and culture of Japan while teaching at the University of Tokyo as a Fulbright Scholar. What she sees as one of her greatest accomplishments, however, was the honor of being one of the founders of *Hypatia: A Journal of Feminist Philosophy,* the premier and oldest journal of feminist philosophy. Most recently, she co-edited the Summer 2009 issue, "Transgender Studies and Feminism: Theory, Politics, and Gendered Realities" with colleague Dr. Talia Bettcher.

In addition to her responsibilities at Cal State L.A., Dr. Garry has served as a visiting professor at UCLA and USC, teaching women's studies and feminist philosophy. Her interests include applied ethics, philosophical method, and epistemology, but she is best known for her work in feminism, which began in the foundational years of the late 1960s. When asked how she felt about being one of the pioneers of feminist philosophy, Dr. Garry responded that the thinking at that time was not, "we're doing something historical." Rather, in step with her direct, no-nonsense style, she recalls that her own thinking as an early feminist was simply, "this needs to be done."

It is this get-it-done attitude that continues to multiply her accomplishments and, coupled with a delightfully humble transparency, enables Dr. Garry to reach students on a personal level as she engages them in invigorating classroom debates. While her own, pithy description is "I know it's messy…" as she hands out the syllabi in her classes, her method of organization emphasizes a commitment to encouraging student exploration of academic themes by keeping their assignment options open. Not surprisingly, Cal State L.A. has recognized her dedication to teaching by honoring her with both the Outstanding Professor Award and the President's Distinguished Professor Award. From her view as a teacher, she expresses appreciation and gratitude for the invaluable cultural perspectives of her students in the richly diverse, multi-cultural atmosphere of Cal State L.A.

A well-known and highly respected leader in her field, Dr. Garry is interested in promoting the integration of feminist philosophy into other curricula She describes the relationship between feminist work and other philosophical fields as "more inclusive than one might think." Two common misperceptions of feminism that she cites are that feminists are anti-male and that feminism is only about white women, and her advice to new feminists is, "to keep their edge—there's still lots of new and important stuff to do."

Dr. Garry will bridge the space between her illustrious academic career and her approaching retirement by continuing to advise M.A. thesis students here at Cal State L.A. She also

looks forward to completing some articles that currently lie on the back burner of her busy life, continuing her activist lifestyle, and spending a semester as the Humphrey Chair of Feminist Philosophy at the University of Waterloo, Ontario in the Fall of 2011.

"I know one thing I'll do, I'll play more tennis," she said in response to questioning about retirement. However, she also confessed to plenty of ideas for new articles—so, this won't be the last we hear from Dr. Ann Garry!

— A. V. B.

Burge's Contextual Theory of Truth and the Liar Paradox[1]

Matthew J. Leonard

1. Introduction

Liar sentences have bothered people for thousands of years—at least since Ancient Greece. However, a noticeably significant number of responses have been given in recent times due to, among other things, Tarski's 1933 paper and the rejection of classical logic (in particular, bivalence) by other writers. The third sentence of Matt Leonard's "Burge's Contextual Theory of Truth and the Liar Paradox" is false. Why is the third sentence of this paper such a problem? Well, suppose that the third sentence is indeed false. If we judge the third sentence false, then it seems to be true, because it (truly) says of itself that it *is* false! Now, suppose that the third sentence is true. If the third sentence is true, then what it *claims* is true. But it claims that it is false; so it must be false! The problem, then, is that we have a sentence, which seems to be both true and false. But everyone (almost everyone, that is) knows that contradictions cannot be true! In fact, according to classical logic (and even many other non-classical logics), we have the consequence that anything trivially follows from a contradiction. So '1 = 0' validly follows from the third sentence of this paper. For this reason alone, the Liar is a significant problem.

What are the *bones* of the problem? What *really* is happening in this dialectic? Not much; and this does not help. If a lot were going on, the chances of there being something to which we could object would be higher. All we need are two (virtually undisputed) rules of inference, the logical law of the substitutivity of identicals, and an instance of a Liar sentence. Given the following two rules of inference,

Semantic Ascent: $\alpha \vdash \mathrm{Tr}(\ulcorner\alpha\urcorner)$[2]

Semantic Descent: $\mathrm{Tr}(\ulcorner\alpha\urcorner) \vdash \alpha$,

and the following instance of the Strengthened Liar,

β β is not true.

we can derive a contradiction. Though the Liar paradox can be generated in a number of ways, the formal proof sometimes runs the following course:

(1)	β = 'β is not true'.	[Given]
(2)	Assume β is true.	[For *Reductio*]
(3)	'β is not true' is true.	[Substitutivity of identicals, (1) and (2)]
(4)	β is not true.	[Descent from (3)]
(5)	β is not true.	[*Reductio* (2)—(4)]
(6)	Assume β is not true.	[For *Reductio*]
(7)	'β is not true' is true.	[Ascent from (6)]
(8)	β is true.	[Substitutivity of identicals, (1) and (7)]
(9)	β is true.	[*Reductio* (6)—(8)]
(10)	β is true and β is not true.	[(5) and (9)]

One relatively recent response to the Liar is provided by those who endorse a contextual approach to semantic paradoxes. In general, contextualism is the view that there is an indexical element involved in the reasoning process of the Liar paradox; given a token of the Liar sentence, the extension of true is contingent upon the context of utterance, and in some theories, the intentions of the speaker. Truth is an indexical notion. If a Liar sentence is not true in some context $\Gamma 1$, then the same Liar sentence could be true in a context $\Gamma 2$, where $\Gamma 2 > \Gamma 1$. There are many different contextu-

alist theories of truth; I will, however, be evaluating only one. In section 3, I will explicate Tyler Burge's (1979) theory. In particular, I will attempt to explain his appeal to conversational implicature and his contention that the extension of the truth predicate varies with shifts in context. In section 4, I will argue that Burge's theory does not succeed. In particular, I will show that there is a crippling dilemma waiting for Burge's contextualist solution to the semantic paradoxes. However, Burge's theory is a somewhat complicated theory. In order to fully see why Burge heads in the direction he does, it is imperative to look back at Tarski's and Kripke's solutions to the Liar—I will do this in section 2. Burge is going to maintain that Tarski's solution is too restrictive and that Kripke's solution fails because of what has come to be known as the 'revenge problem.' It is necessary to briefly review Tarski's and Kripke's theories of truth because the plausible reasons for rejecting them form the main support for Burge's contextual solution; but also, they are embedded in the objection I raise against Burge's theory near the end of the paper. To this brief background then, I now turn.

2. Tarski's Hierarchical Solution and Kripke's Paracomplete Solution

Tarski maintained that the threat of paradox occurs when the truth predicate for a language L_1 resides in L_1 itself. This is why natural language generates paradox. Thus, he proposed that the truth predicate for a language L_1 must be placed in a metalanguage L_2. If we start with an interpreted language L_1, which excludes a truth predicate, we can then add a truth predicate to form L_2 and make claims regarding the veracity of sentences in L_1. For instance, for a sentence φ in L_1, when can claim in L_2 that 'φ is true.' The hierarchy is infinite. For any sentence φ and any level n, we can only claim that 'φ is true' in L_{n+1}. How does this solve the Liar paradox? It solves the Liar paradox because it blocks the formulation of a Liar sentence. Since there are no Liar sentences, there is no paradox.

Though Tarski (1933) did successfully block the Liar in giving his definition of truth, most people (I think rightly) want to say that while Tarski's definitions of truth and designation are fruitful for metalogical and metamathematical theories, they are too restrictive for our *ordinary* notions of truth and meaning. One's theory of the semantic paradoxes should match our pre-theoretic intuitions about natural language, rather than 'block' paradoxical sentences in an artificial language. Hence, Tarski's solution to the Liar is too restrictive.[3]

A more recent solution is the paracomplete solution to the semantic paradoxes. Paracomplete solutions maintain that β does not have a truth-value (it lacks truth conditions). Perhaps the most influential response endorsing truth-value gaps is Kripke's (1975) theory of truth. He claims that the set of true sentences must be a 'fixed point' of some monotonic evaluation scheme (where he uses Kleene's strong three-valued logic). Kripke elegantly constructs his theory in the following way.[4] Kripke begins with a language L_1 that lacks a truth predicate. L_1 itself has a classical valuation scheme. Kripke then extends L_1 to L_2, which contains a truth predicate. However (unlike Tarski's theory), the truth predicate can be applied to every sentence of L_2 (including all of the sentences of L_1). L_1 can be interpreted with a classical model M_1. Kripke proposes to build up a model M_2 for the expanded L_2. The truth predicate has both an extension and an anti-extension. The extension is the set of true sentences and the anti-extension is the set of false sentences; though these are disjoint sets, they do not exhaust the entire domain. Kripke employs an inductive method here. Start with an empty extension and an empty anti-extension. Start throwing in true sentences to the extension and false sentences into the anti-extension. Liar sentences do not appear in either set, and thus are viewed as 'gappy.'[5]

This common sort of response to the Liar, however, has been met with a serious problem. It is often referred to as the 'Revenge of the Liar,' or 'Strengthened Liar reasoning.' The revenge problem is not *really* a new problem; it is simply another instance of the Liar masked for truth-value gap responses to the original

Liar. Technically, the original Liar is of the following form,

β_{OL} β_{OL} is false.

When met with the original Liar, one can just claim that β_{OL} cannot be true and it cannot be false; 'No problem, β_{OL} lacks a truth value.' However, consider again an instance of the Strengthened Liar,[6]

β_{SL} β_{SL} is not true.

The Strengthened Liar is supposed to show that β_{SL} cannot be true, cannot not be true, and *cannot not have a truth-value*. If β (from now on, just assume that β has the strengthened form) is neither true nor not true, then in particular it is not true. But if it is not true, then it seems that β is not true (since that is what β seems to tell us). Therefore, β seems to be *true* in an important sense; β is true 'after all'![7] The same revenge problem can be applied to theories that state that β is meaningless or vacuous. If one claims that β is meaningless, then it seems that β would not be true; from this, we can generate a contradiction. Likewise, if one responds by claiming that β is vacuous, i.e., neither true nor false, then β in particular is not true. From this we can generate a contradiction as well. Another intuitive response might be to claim that β is empty and simply does not say anything. I think there are two problems with this type of response. First, sentences can non-problematically predicate things of themselves, as in γ,

γ γ is composed of six words.

It seems that β is non-problematically predicating something of itself, just as γ is. Second, though some statements do lack truth conditions (for example, it is clear that interrogative sentences do, and perhaps sentences with vague predicates do too), it seems that the Strengthened Liar shows that β cannot be one of these instances. Once one asserts that β lacks truth conditions, one gets caught in its web (unlike interrogative sentences or sentences with vague predicates, because there is nothing paradoxical about them).[8] In any event, let us assume that this is right, since this is

an essential factor that drives Burge to his solution to the Liar.

3. A Précis of Burge's Theory

Burge wants to distance himself from both the Tarskian and Kripkean solutions to the Liar. He rejects the former for the same reasons most people do, as I've mentioned above (i.e., it is too restrictive for our ordinary notion of truth, and so on). He rejects the latter, i.e., truth-value gap theories, because of the revenge problem. As a result, he posits a hierarchical theory that, though similar in some respect to Tarski's, differs by attempting to meet some of the pre-theoretic semantic intuitions Tarski's theory did not account for. In particular, he does this by claiming that our *ordinary* notion of truth is indexical and that the extension of our ordinary notion of 'true' shifts from context to context, as opposed to claiming that some technical notion of truth ('true at level L', for instance) is indexical. In "The Concept of Truth in Formalized Languages," Tarski sought to block the Liar by assuming that he was dealing with some purely extensional concept of truth, not the ordinary notion. In fact, he argued that natural language was inconsistent and inevitably generated the Liar. Burge, on the other hand, wants to treat the Liar as if it *is* utilizing our ordinary notion of truth. In this section, I will offer a brief summary of his theory.

Burge argues that there is a hidden conversational implicature and a shift in extension (parallel with a shift of context) that occurs in Strengthened Liar reasoning. According to Burge, Strengthened Liar reasoning runs the following course:

Step 1. An occurrence of a Liar like sentence.

Step 2. The Liar sentence is not true.

Step 3. The Liar sentence is true after all.

Most solutions to the Liar have either ignored such reasoning or attempted to block it by formal means. Burge, on the other hand, thinks a more satisfying approach is to interpret the reasoning so as to justify it. He thus takes the Strengthened Liar as a model for

how we *should* think when confronted with the semantic paradoxes. Consider the following very plausible scenario (and notice the corresponding Steps 1–3):

> Suppose I think both that I am in 398 DODD and that the professor, at this moment, in 399 DODD is a fraud. So I write on the board at 4:00P.M. on 5/21/09, (Step 1) 'There is no sentence written on the board in 399 DODD at 4:00P.M. on 5/21/09 which is true as standardly construed.' However, unbeknownst to me, I am in fact the one in 399 DODD, and this is the only sentence written on the board. The usual reasoning shows that this cannot have truth conditions; thus, it is not true. (Step 2) So there is no sentence written on the board in 399 DODD at 4:00P.M. on 5/21/09 which is true as standardly construed. But we have just stated the sentence in question. (Step 3) Thus, it is true after all.

In Steps 1 through 3, Burge argues that there does not seem to be a grammatical or semantical change in the expressions involved. If this is true, then the shifts in evaluation should be explained pragmatically. However, in Steps 2 and 3 we seem to predicate both truth and non-truth of the same sentence. Pragmatics cannot account for this. Like most people, Burge does not want to accept the conjunction of Step 2 and Step 3, i.e., the dialetheism of philosophers like Graham Priest. Thus, Burge concludes that there must be a shift in extension for the predicate 'is true.' Burge argues that the conversational implicature occurs between Steps 1 and 2, and the shift in extension (and likewise, a shift of context) occurs between Steps 2 and 3. Let me briefly explain these Steps.

For Steps 1 and 2, if there is no grammatical (or semantic) changes in meaning, nor is there a shift in extension of the truth predicate, the change must be explained pragmatically. Burge attempts to explicate this pragmatic change in terms of implicature. He proposes the following implicature: "*sentences being referred to or quantified over are to be evaluated with the truth schema for the occurrence of 'true' in the evaluating sentence*" (Burge 1979,

p. 95). The implicature is manifested in the following way. In Step 1 the implicature is that the sentence quantified over should be evaluated with a truth schema containing true_i.[9] However, the sentence turns out to be the problem sentence itself, and accepting this implicature engenders a contradiction. In Step 2, the implicature of Step 1 has been canceled. The problem sentence as it occurs in Step 2 is not true_i (in the same way that it was not true_i is Step 1), but it no longer has truth_i conditions. This leads to the next move.

Though there is not a shift in extension in Steps 1 and 2, there must be a shift in extension *somewhere* to account for the fact that both *truth* and *non-truth* are predicated of the Liar. This is the most appropriate place for such a shift to occur. The shift from Step 2 to 3 involves a transition from a sentence that is not true to a sentence that is true—and this is the same sentence! Burge explains this move by regarding semantical predicates as indexical, and provides a formal system intended to model natural language. He gives both a formal and a pragmatic explanation of his theory. The former explicates the structural *relations* between the extensions of different occurrences of 'true,' and the latter explicates *how* the extensions of indexical occurrences are established in context. I will briefly discuss each in turn.

Burge wants to stipulate a formal system (which matches English) that defines a pathological_i sentence, as interpreted in a context.[10] Burge stipulates that *pathologicality* is a disposition to produce disease for certain semantical evaluations. Thus, the Liar comes out pathological.[11] *Rootedness* is defined as the lack of pathologicality, i.e., a formula's being *rooted* means that it is nonpathological, and (roughly) that it has a truth-value.[12] Burge then distinguishes extensions of 'true' by marking occurrences of them with subscripts beginning with i. In the Strengthened Liar case, for Step 2 Burge claims that the Liar sentence is not true—he marks this initial context of utterance 'true_i.' In Step 3, and from a broader application of truth, he claims that the Liar is true—Burge calls this context of utterance 'true_k,' where $k > i$. He argues that though pathological_i sentences are not true_i, pathologi-

cal$_i$ sentences are nonpathological$_k$, and thus true$_k$. All rootless$_i$ sentences are not true$_i$. So a sentence and its negation may both not be true$_i$, though one or the other will be true$_k$. Thus, Burge offers the following restricted Tarskian truth schema:

(T) ($\forall i$) If a sentence ⌜φ⌝ is rooted$_i$, then ⌜φ⌝ is true$_i$ iff p.

where ⌜φ⌝ names any well-formed sentence in Burge's system and 'p' is the sentence itself. The pressing question remains: how does any of this work for English? How do these hierarchical subscripts get fixed? This leads to the pragmatic explanation of his theory.

Burge provides the following plausible pragmatic rules for determining how the subscripts are established in context.[13]

> *Justice* = Subscripts should not be assigned so as to count any given sentence substitutable in a truth schema instead of another, without some reason.
>
> *Verity* = Subscripts on occurrences of 'true' are assigned so as to maximize the acceptability of truth schemas to sentences and minimize attributions of rootlessness.
>
> *Minimalization (Beauty)* = The subscript on occurrences of the predicate 'true' is the lowest subscript compatible with the other pragmatic principles.

First, notice how the formal and pragmatic principles apply to the Strengthened Liar (notice the brackets indented to the right):

Step 1. β β is not true. [i.e., this is just a Liar token.]

Step 2. The Liar sentence is not true. [i.e., β is not true$_i$.]

Step 3. The Liar sentence is true after all.
[i.e., Step 2 is true$_k$.]

For a more complicated version, consider how Burge's formal and pragmatic principles apply to Kripke's version of the Deferred Liar. Suppose Dean utters,

(D) All Nixon's utterances about Watergate are untrue.

And then suppose Nixon utters,

(N) Everything Dean utters about Watergate is untrue.

By the principle of Justice, each person's truth predicate should be assigned the same subscript, *i*. By the principle of Verity, *i* must be high enough to interpret any statement by Dean or Nixon other than (D) or (N). By the principle of Minimalization, *i* must be no higher. If Dean uttered at least one truth_i about Watergate, then Nixon's (N) is rooted_i and not true_i. If none of Nixon's utterances other than (N) are true_i, then, since (N) is not true_i, Dean's (D) is true_i. If Nixon did say something true_i about Watergate other than (N), then Dean's (D) is rooted_i but not true_i. Now suppose that no utterance about Watergate other than (D) by Dean is true_i. If none of Nixon's utterances other than (N) are true_i, then neither utterance is rooted_i and both vacuously not true_i. If at least one of Nixon's utterances other than (N) is true_i, then Dean's utterance is rooted_i and not true_i, and Nixon's utterance is rooted_i and true_i.[14]

4 A Dilemma for Burge's Contextual Theory of Truth

The initial appeal of contextual approaches to the semantic paradoxes is that they accord with some of our intuitions about how truth works, and in particular, how Strengthened Liar reasoning should proceed. As enticing as this appeal might be, there is a worry with Burge's contextual theory that throws doubt on whether this type of response to the Liar is, in fact, the right type of response. I want to mention both a general worry for all contextualist solutions to the Liar, and a specific worry with Burge's theory. The specific worry is just a problem with Burge's response to the general worry; so first, let me mention the general worry. The general threat to contextualism emerges when the Strengthened Liar is reformulated in a way that purposely refers to hierarchical contexts; this formulation is sometimes referred to as the Super-Liar. What type of response can the contextualist provide for sentences like 'This sentence is not true at any level, or in any

context,' or sentences like ψ?[15]

ψ $(\forall i)$ ψ is not true$_i$

It seems that contextualism faces the same sort of paradox Tarski, Kripke, and earlier theories faced. Either ψ is not true$_i$ at any level i or ψ is true at some level n. Suppose it is not true$_i$ at any level. But that is just what ψ says of itself. Hence, ψ is true$_k$, where $k > i$ (i.e., ψ is true 'after all'). On the other hand, suppose ψ is true at some level n. If that is the case, then ψ should come out false at n, because it says of itself that it is not true at any level. In both cases, contextualism seems to be unable to account for ψ.

As far as contextualism goes, I think this formulation counts as a serious problem. Contextualism seemed most plausible when showing how it circumvented the revenge problem; the conceivability of Super-Liars, at least *prima facie*, seems to immediately discard the alleged solution contextualism proposed. Some philosophers think that this is a knock-down argument against contextualism.[16] Whether or not this is so, is not the point of this paper. What I want to note is that Burge already knows that this version of the Liar can be generated against contextualism; in fact, he provides a response in advance in his original paper. What is puzzling is that no one seems to address his response. They just assume that the Super-Liar succeeds in offering a problem for contextualism. Perhaps it does succeed; nonetheless, it will not satisfy Burge because it does not address his antecedent response. However, I'm going to argue that the really devastating problem emerges when we put pressure on Burge's response.

Foreseeing the general potential problem, Burge writes, "Attempts to produce a 'Super Liar' parasitic on our symbolism tend to betray a misunderstanding of the point of our account. For example, one might suggest a sentence like (a), '(a) is not true at any level'. But this is not an English reading of any sentence in our formalization. Our theory is a theory of 'true', not 'true at a level'" (Burge 1979, p. 108). Clearly, Super-Liars can be generated using the technical notion of 'truth at a level' or other technical ingenuities.[17] It seems, then, what Burge is hoping to do is show that they

cannot occur with our ordinary truth predicate, and thus, they pose no real threat for his account of the semantic paradoxes. Imagine me in the same classroom scenario above with the same false beliefs. Suppose I then assert, 'There is no sentence written on the board in 399 DODD at 4:00 P.M. on 5/21/09 which is true, in any context, under any level of evaluation.' It is obviously conceivable that I assert such a statement in such a scenario. But Burge's point seems to be that it would not be ordinary English. Philosophers might assert such a statement, but ordinary folks wouldn't. It seems to be very technical. I think this is right to an extent. Ordinary language users would probably not find themselves uttering such a statement. I do not think, however, that this completely removes Burge's theory from objection. Something still seems to be wrong, and to that I now turn.

I am going to argue that it is not inconceivable to construe a Super-Liar in ordinary English. Some formulations are obviously not formulations in ordinary language. Suppose Ralph is walking down the street and shouts the following to a crowd of people, "To the sentence immediately following this one, I stipulate the name 'ψ_1'. 'ψ_1' is not Tarski-true in L, in any of the transfinite metalanguages of L." The crowd would completely ignore Ralph. Why? Because his utterance is not ordinary English. Suppose further, that a philosopher was in the crowd. Perhaps she might chuckle for a few seconds, but she would immediately know that Ralph's sentence was not ordinary English.

However, I want to maintain that Super-Liars can still be generated in ordinary English. There are two characteristics of ordinary language (at least, in English) that support my claim. The first characteristic of natural language is a *contextual* characteristic. The examples are plentiful. John says to his wife Mary, 'How much can I drink tonight?', to which Mary replies 'Well, it depends on the *context'* (are you going to a football game, or are you going to Mass?). Someone makes a political remark on TV, and Jones says to his friend, "That's a really narrow way of looking at things" (where the 'narrow way' here is some sort of context). This contextual characteristic of natural language shows

that ordinary language users understand that language is somehow broken-up, or contextual, or (perhaps even) hierarchical. The second characteristic of natural language is a *modal* characteristic. The examples, likewise, are plentiful. Ordinary folk say the following sorts of things all the time, "Well, what you said would have been true if such-and-such (if McCain/Palin would have won the 2008 ticket)," or "Well, what you said might have not been true if such-and-such (if the Steelers would have won the 2009 Super Bowl)." What I think this second characteristic shows is that normal (non-philosopher) language speakers sometimes take a sentence and compute its truth-value under different circumstances.

One might quickly respond in the following way: 'Sure, these two characteristics are genuine characteristics of natural language; but that doesn't show that ordinary speakers understand, or use, the technical notion of truth-at-a-level required to generate a Super-Liar.' I agree with the first clause; this isn't intended as proof that natural language speakers understand the technical notion of truth. I'm not claiming this; nor am I claiming that it is necessary to generate a paradox. These two characteristics are intended to show that normal English speakers might on occasion utter the following statements, which seem to utilize our ordinary truth predicate (not the technical one Tarski constructed),

ψ_2 What I'm saying will never be true, in any context.

ψ_3 What I'm saying could never be true.

The problem, then, is that ψ_2 and ψ_3 seem to clearly be normal English sentences, capable of being uttered by normal English speakers. Burge's theory, though it does turn out consistent for the Strengthened Liar and the Dean/Nixon Deferred Liar, turns out inconsistent for ψ_2 and ψ_3. For ψ_2 and ψ_3, Burge's theory would land in paradox for the same reasons it did when applied to the more formal ψ,

ψ $(\forall i)$ ψ is not true$_i$

Namely, either ψ_2 is not true$_i$ at any level *i* or ψ_2 is true at some level *n*. Assume the former, and then ψ_2 will be true at level k, where $k > i$ (i.e., ψ is true 'after all'). Assume the latter, and ψ_2 should come out false at *n*. Either way, we have a paradox. The same reasoning could apply to ψ_3.

What I've attempted to show, then, is that it is not completely inconceivable for one of these Super-Liar instances to occur in natural language. In other words, it is possible for natural language to be such that one of these instances is stated in them. I don't need to show that natural language users completely understand how to evaluate Super-Liars, in the same way that natural language users don't need to be able to understand how to evaluate instances of the Strengthened Liar (as in the DODD classroom case mentioned above). But just as the Strengthened Liar can be uttered in natural language, so too can (some) instances of the Super-Liar. Thus, Burge's theory must be able to apply to it as well.

One plausible way Burge might want to respond is to argue the following: *Just as Tarski's proposal allowed us to talk about a sentence* φ *in a language L only in a corresponding metalanguage L*, so too can we only talk about a sentence* ψ *in a context* σ *by indirectly referring to a larger context* σ**. But if this is the case, then we wouldn't even be able to formulate a Super-Liar instance because whenever a sentence referred to its own context, it would slightly shift to a greater context*.[18] This would indeed block the formulation of a Super-Liar instance. However, I think this sort of reply won't work. One of the central criticisms of Tarski's approach (that Burge himself mentions) is that in light of naturally occurring cases of self-reference, his ruling the Liar sentence as syntactically 'ill-formed' seems to be overkill. Though his solution works (or blocks the Liar) for formal languages, his solution seems implausible as applied to many naturally occurring uses of our ordinary truth predicate. I think I can make the same sort of reply to the attempted blocking of the Super-Liar; namely, that this response seems to be, analogously, too restrictive. Though it blocks the *technical* formulation of a Super-Liar, it seems that in natural language, when ordinary language users talk about

contexts, they don't always refer to them in such a technical way. Imagine someone asserting something like, "Torturing the innocent is enjoyable," to which someone replies, "That is morally reprehensible in any context," or "That is wrong in any circumstance." When one makes a universal claim about *all* contexts in natural language, it doesn't seem that she is 'stepping back' to a greater context (beyond all the contexts to which she is referring). And so, I can say that it seems a bit restrictive to try to block Super-Liars in natural language, just as Tarski's blocking the Liar was too restrictive because it didn't account for Liar cases in ordinary English. It seems that it is plausible for a Super-Liar to occur with our ordinary truth predicate, and so one's solution of the semantic paradoxes should be able to apply to it as well.

From the beginning of the paper, I have indirectly juxtaposed Burge's solution to the Liar with Tarski's solution. Though they are strikingly similar in the way of appealing to hierarchy, they are strangely distinct with respect to the specific notion of truth to which their theories are targeted. Burge argued that Tarskian and Kripkean theories wrongly conceive of truth in Liar instances as some technical notion ('true at a level,' 'true in L,' 'Tarski-true,' or something). He wants to argue that our *ordinary* notion of truth is indexical and that it shifts in extension in Strengthened Liar reasoning—not that 'true in L' is indexical and shifts in extension. Kripke's paracomplete solution, utilizing some technical notion of truth, was unable to account for the revenge problem; thus, contextualism seemed more plausible because it *could* account for the revenge problem, as it occurred in ordinary language. Burge, however, wants to provide a solution to the semantic paradoxes as they occur in genuinely empirical (colloquial) circumstances. Thus, he wants his theory to be able to account for paradoxes as they occur in the DODD scenario and the Nixon/Dean Deferred Liar scenario.

For a theory of truth, one (ideal) component I would like it to have is for it to treat the semantic paradoxes as *monolithic*, in the sense that they are to be evaluated similarly. I have a strong inclination to say that all instances of the semantic paradoxes utilize

the same sort of truth. Perhaps this is too strict of a demand. For that reason, I haven't attempted to prove this in this paper; but more importantly, I don't need to. I have at least shown that the Strengthened Liar and the Super-Liar can both have instances in natural language, uttered or written by normal English speakers. This is devastating for Burge's theory, because as a consequence, we have the result that the Strengthened Liar and the Super-Liar utilize the same notion of truth. Hence, Burge's proposed bifurcation between the two sorts of Liars in question doesn't hold.

To be honest, I'm not entirely convinced as to whether one should treat the Strengthened Liar as utilizing our ordinary notion of truth or some technical notion of truth. To explore this would require another paper. However, what I have shown above, is that the *same sort of truth* is utilized in both kinds of Liars, i.e., they either both utilize our ordinary notion of truth or both utilize a technical notion of truth. They seem to be rather similar; both can be formulated in natural language. Perhaps there are other reasons as to why we should think of the Strengthened Liar as utilizing one type of truth and the Super-Liar the other; but what is clear is that Burge's initial reasons for thinking of the Strengthened Liar as uniquely utilizing one sort no longer apply. Since both the Strengthened Liar and the Super-Liar seem to utilize the same sort of truth, Burge's theory falls in the following dilemma:

(1) the Super-Liar utilizes truth at a level, or

(2) the Super-Liar does not utilize truth at a level (and instead utilizes the ordinary notion of truth).

If the former is correct, then the Strengthened Liar utilizes some technical notion of truth as well. But the initial plausibility of Burge's position was due to the fact that it could explain the Strengthened Liar in terms of our ordinary notion of truth. If we choose this horn, then the initial reasons for Burge's theory seem to be significantly weakened. If the latter is correct, then Burge owes an account of how his theory deals with the Super-Liar (i.e., 'What I'm saying will never be true, in any context'). Either way,

Burge's theory seems to be in a serious predicament.

Notes

1. I would like to thank the audience at the 2010 Berkeley/Stanford/Davis Philosophy Conference at UC Berkeley for the many comments provided, especially Robbie Hirsch, who provided a very helpful formal response to the paper. I would also like to thank the following people for comments: Michael Hatcher, Michael Anderson, Zeph Scotti, and Prof. David Pitt. I would especially like to thank Prof. Mark Balaguer for offering comments on numerous drafts of the paper.

2. Where $\alpha \vdash \mathrm{Tr}(\ulcorner\alpha\urcorner)$ is read, 'From α, you may validly infer $\mathrm{Tr}(\ulcorner\alpha\urcorner)$,' and where $\ulcorner\alpha\urcorner$ refers to the Gödel number of α. The debate concerning the nature of truth is obviously much too large to take on in this paper, but almost everybody accepts these two rules of inference. Two more rules of inference almost everybody accepts as well are

   Negative Semantic Ascent: $\neg\, \alpha \vdash \neg\, \mathrm{Tr}(\ulcorner\alpha\urcorner)$

   Negative Semantic Descent: $\neg\, \mathrm{Tr}(\ulcorner\alpha\urcorner) \vdash \neg\, \alpha$,

   I'll only be using the first two rules of inference in this paper. Further, I am using corner quotes, specifically, to quantify over metalinguistic variables.

3. See his "The Concept of Truth in Formalized Languages." Reprinted in his *Logic, Semantics, Metamathematics*. Oxford: Oxford University Press, 1956, pgs. 152–278. Originally published in 1933.

4. The following brief exposition is roughly taken from Beall and Glanzberg (forthcoming).

5. See his "Outline of a Theory of Truth" *Journal of Philosophy* 72 (1975): 690–716.

6. The terminology here is van Fraasen's.

7. The Strengthened Liar shows the problem in a more intuitive way than the original Liar. I think you can still generate a 'revenge problem' from the original Liar, but I think it is harder to see. It goes like this (I'm getting this from Burge (1979)): Suppose β_{OL} is not true and not false, i.e., it is truth-value-less. β_{OL} would then be 'not false.' But "β_{OL}' is not false' is the negation of β_{OL}, so the negation of β_{OL} is true. But if the negation of β_{OL} is true, then β_{OL} itself is false. But then, that is precisely what β_{OL} says of itself. So there is a paradox.

8. There is an entire volume devoted to this problem; see JC Beall, *Revenge of the Liar: New Essays on the Paradox*, Oxford: Oxford University Press, 2008.

9. Where the subscript *i* simply assigns a context to 'true'. I will make this more clear, shortly.

10. He provides three elegant constructions of such a theory, preferring the last. I will not include the axioms of these constructions; rather, I will mention the basics of his theory.
11. The Truth Teller (i.e., a sentence which says of itself that it is true) will also come out pathological.
12. Rootedness is essentially the same notion as groundedness in Kripke's theory.
13. I'll mention the third pragmatic rule as well, even though it was not introduced until his later paper, "The Liar Paradox: Tangles and Chains" (and his 1982 postscript to "Semantical Paradoxes").
14. See Burge, "Semantical Paradox", 110, and Burge, "The Liar Paradox", 360–361.
15. This formulation and the following brief discussion of it are a basic version I came up with, primarily inspired by Cory Juhl's (1997) paper. More complicated formulations are discussed by Juhl (1997) and Gauker (2006). For simplicity, I will apply ψ to Burge's theory, though it can be applied to all contextual solutions to the Liar.
16. Juhl (1997) and Gauker (2006), for instance.
17. Juhl (1997), for instance, offers the following context-sensitive Liar case:

    φ φ is not true$_i$ at any level i up through F(c, φ),

    where F(c, x) returns the least level (ordinal) α such that, x is true$_\alpha$ in context c, where c is f(x), the least context in which x is uttered. Therefore, in context c, we have the following dilemma; either

    > φ is not-true$_i$ at every level i, in which it ought to be true at some level (i.e., F(c, φ)+1), or
    >
    > φ is true at some level, and F(c, φ) specifies the least such level. In that case, φ ought to come out false at that level, since it says that there is no such level up through F(c, φ).

18. This objection was given to me by Robbie Hirsch.

References

Beall, JC, ed. (2007) *Revenge of the Liar: New Essays on the Paradox* (Oxford: Oxford University Press)

Beall, JC and Michael Glanzberg (Forthcoming) "The Liar Paradox" *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta, ed.

Burge, Tyler. (1979) "Semantical paradox," *The Journal of Philosophy* 76, pp. 169–198. Reprinted in Martin (1984)

Burge, Tyler. (1982) "The liar paradox: tangles and chains," *Philosophical Studies* 41, pp. 353–366

Gauker, Christopher. (2006) “Against stepping back: a critique of contextualist approaches to the semantic paradoxes,” *Journal of Philosophical Logic* 35, pp. 393–422

Juhl, Cory. (1997) “A context-sensitive liar,” *Analysis* 57, no. 3, pp. 202–204

Kripke, Saul. (1975) “Outline of a theory of truth,” *The Journal of Philosophy* 72, pp. 690–716. Reprinted in Martin 1984

Martin, Robert L, ed. (1984) *Recent Essays on Truth and the Liar Paradox* (Oxford: Oxford University Press)

Tarski, Alfred. (1933) “The concept of truth in formalized languages,” Reprinted in: A. Tarski (1956) *Logic, Semantics, Metamathematics* (Oxford: Oxford University Press), pp. 152–278

Identifying Cognitive Phenomenology in Husserl: Making Sense of Demonstratives

Ali Mohsen

It is a commonly held belief within analytic philosophy that thoughts have intentionality, but no phenomenology. David Pitt challenges this assertion and claims that thought content is a purely cognitive phenomenology, separate from all other sorts of phenomenologies. However radical this may seem for mainstream analytic philosophy, Pitt's view is not all too different from Husserl's. The difficulty in making this comparison arises with cases of demonstrative thoughts, i.e., thoughts that use a demonstrative.[1] Pitt maintains that the content of a demonstrative in thought is something corresponding to its linguistic character, and that it refers to an object relative to a context or worldly situation. I will demonstrate that aside from obvious differences in exposition, Pitt's theory is also found in Dagfinn Føllesdal's "west coast" interpretation of Husserl. Even though Husserl never says that thoughts have their own cognitive phenomenology, interpreting him in this way is compatible with his theory of perceptual acquaintance and makes the most sense of his theory of demonstratives. Husserl is not explicit in his discussion of indexicals and demonstrative reference, but I believe that by refining his theory of perceptual acquaintance with the role of cognitive phenomenology in mind, we shall clarify and enhance the "west coast" interpretation.

Pitt makes three claims about the phenomenology of thought: "…what it is like consciously to think a particular thought is (1) different from what it is like to be in any other sort of conscious mental state (i.e., *proprietary*) and (2) different from what it is

like consciously to think any other thought (i.e., *distinctive*)," and, "…(3) the phenomenology of a thought *constitutes* its *representational* content (i.e., is *individuative)*." He sums these claims up in a single principle (P), "Each type of conscious thought—each state of consciously thinking that p, for all thinkable contents of p—has a proprietary, distinctive, individuative phenomenology." (Pitt *The Phenomenology of Cognition* 2004, pp. 4–5). Pitt says that if (P) were not true, then we would not be able to (a) distinguish between our conscious thoughts and those of others, (b) distinguish between different thoughts within our own consciousness, nor (c) identify our thoughts as the thoughts that they are (ibid., p. 7). Since we would not be able to introspect a conscious thought unless it had a particular phenomenal character, there must be an introspective experience of what it's like to think a thought, a distinctly cognitive phenomenology (ibid., p. 10).

On this account, to introspect a conscious thought is to be attentive to an internal experience; this is not to be confused with perception in the ordinary sense, i.e., attending to external objects. A consciously introspected mental object is a part of the overall introspective experience, whereas the external object of a perception is distinct from the perceptual experience. Moreover, simple introspections of mental objects are not belief or knowledge states; rather, they are recognitions of the qualitative properties of the mental objects to which we are attended (ibid., p. 10). To have introspective knowledge about the thought that p and to identify which thought it is, is to grasp that the thought in question *is* the very thought that p, i.e., to understand the content of the thought as being that p. But we can only have introspective knowledge of the thought that p in virtue of thoughts having their own peculiar phenomenological properties. Thus, we identify thoughts via introspection of their unique phenomenal properties (ibid., p. 11).

With respect to demonstrative thoughts, Pitt says that they refer to objects relative to a context of utterance. The content of a demonstrative in thought is just its linguistic character, and they cannot refer to objects without the aid of a context. However, one may argue that the content of demonstratives are referent-depen-

dent, i.e., they are partially, if not entirely, determined by their referents. If this is the case, then demonstratives in thought would not be individuated in virtue of their intrinsic phenomenal properties, and thus their contents could be comprised of other types of phenomenologies or even non-phenomenal entities, e.g., percepts or physical objects (Pitt *Demonstrative Thoughts*, pp. 1–2).

Pitt offers three possible reasons to support the intuition that demonstratives are referent-dependent and argues against them. The first example that he gives is a scenario in which two people each think ^{T}I'm hungry^{T}.[2] A common intuition is to say that these are two different thoughts because they are about different people, or have two different referents. Since the thought's meaning determines reference and the two thoughts have different referents, it would follow that the thoughts have different meanings. But this opposes Pitt's view that demonstratives are individuated by their intrinsic phenomenal properties; he would say that they are in fact the same thought (ibid., pp. 2–3). He resists this intuition because he believes that thoughts are mental states, and that the truth conditions of mental states do not influence their contents. The ^{T}I^{T} components in the two thoughts are tokens of the same type and determine their referents relative to context (ibid., pp. 5–7).

The second objection that Pitt presents is that demonstrative thoughts are not individuated by their intrinsic phenomenal properties, but by their evaluations at other possible worlds. On this intuition, if we evaluate the content of a demonstrative thought at another possible world, it is the demonstrative's referent that is being evaluated, not its linguistic character. Further, thoughts could not be phenomenologically constituted, since the evaluation of thoughts in possible worlds concerns referents and truth conditions, not the phenomenal content of thoughts (ibid., pp. 3–4). However, Pitt's view is that ^{T}I^{T} does not individuate anyone in particular; it is a general term by virtue of its linguistic character and usage. Possible worlds invoke a referent, but demonstratives, by their very construct, are not specific enough to determine a referent without a context. Thus, demonstratives are non-rigid and their extensions can vary at different worlds (ibid., pp. 10–13).

This is just the peculiar nature of demonstratives and this is not at odds with the position that thoughts are phenomenally constituted.

Lastly, it may be the case that if we do not know the referent of the demonstrative in thought, then we do not understand the thought. For example, if I think ^{T}That's a strange thing^{T} in a dark room with nothing in particular in mind, then intuitively, it may seem as if I have not understood what I thought, or perhaps not succeeded in thinking anything at all (ibid., p. 4). Pitt responds by saying that understanding is attained in degrees. Even though the thought mentioned above may be "empty", it still has meaning and can be understood. We know from this thought that it purportedly refers to an object, and that the object has the quality of being strange. We would have a better understanding of the thought if the context was rich enough to provide a referent, but we still *understand* what the thought entails (ibid., pp. 13–14).

The difficulty in comparing Pitt and Husserl's views of demonstratives is the complexity of Husserl's account of perception. Husserl's explanation entertains multiple phenomenologies in one loaded experience, whereas Pitt gives a separate account of each phenomenal experience. However, both views allow for different phenomenologies to work side by side. Husserl makes a distinction between an act's object (referent) and its meaning.[3] He defines the term "noema" as an abstract entity that bears the meaning of an act. We do not directly perceive the noema; rather, the noema is a sort of function that picks out the object of thought and it carries intentionality insofar as it prescribes objects for acts.[4] The noema is immanent within an act and we can become aware of its presence through reflection. We can recognize that our thoughts purportedly refer to things, no matter if the object is existent, non-existent, determined, or partially determined (Husserl *Ideas* 1976, §88–89).

Husserl distinguishes between an act's object and noema in order to solve the problems that the concept of intentionality faces, i.e., to explain how it is that we can refer to non-existent objects and how different acts can refer to the same object in different ways.[5] In the first case, I could desire to catch a lepre-

chaun, even though leprechauns do not exist. Husserl would say that my act has no object and that the act is only comprised of its content, or "intentional experience". Regarding the second problem, I could think of my dining table as 'my dining table' while my friend could think of it as 'the table that wobbles'. Both thoughts refer to the same object, but the noemata are different insofar as the thoughts refer to my table with a different mode of presentation. Conversely, thoughts with the same noema refer to the same object, e.g., when my friend and I both refer to my dining table as 'the table that wobbles', the noemata of our acts are the same[6] (ibid., §99).

A crucial distinction pertaining to the noema in perceptual experiences is the difference between what Husserl calls the fulfilling sense and the empty sense. Husserl uses this distinction to explain how we acquire knowledge in degrees or stages. A perceptual experience achieves two things: it presents the subject with an object as having characteristics and predicates, and it forms an intentional relation with the object. An act with a fulfilled sense is supported by perceptual evidence and the existential veridity of objects is gained in degrees with respect to the fullness of the fulfilling sense. Thus, if a fulfilled sense regarding the existence of an object were also corroborated with the testimony of other's, then the likelihood of the object's existence would greatly increase. The empty sense, however, is created in imagination due to the lack of direct perceptual acquaintance with an object. Thoughts that do not refer to anything in particular have phenomenal content, but their senses are empty. Similarly, thoughts of non-existent or non-present objects would also have an empty sense, although past experiences and a richer context may present them as slightly fulfilled (Dreyfus 1982, pp. 97–103).

Let us imagine that unbeknownst to me, an aircraft will soon fly overhead and land nearby. While the sky is clear, any thoughts of mine about an aircraft flying nearby would be empty since I would not be perceptually acquainted with an aircraft. Shortly after this, I see an object in the sky flying high above me. My thoughts about the object high up in the sky would be slightly fulfilled; I

would be referring to an object in the sky, but I would not know its identity. If, however, the object flies lower to the ground and I see and hear that it is an airplane, my thoughts about the object would become much more informative. The perception of an airplane is veridical insofar as it is added to the context of my experience and allows me to identify it. Let us imagine further that I was mistaken about the airplane: what if the object landed in front of me and I realized that it is actually a helicopter? Even though I was wrong about what sort of aircraft was flying in the sky, this does not mean that I was referring to a different object. As Husserl would say, my noema "exploded" insofar as my horizon of expectations was grounded on a false assumption about the nature of the object. In light of new circumstantial evidence I can construct a new noema that is more fulfilled due to more accurate perceptual acquaintance with the object. Regardless of my false assumption, my thoughts referred to the same aircraft as that which landed; and through a more informative context of experiences, I was able to correct my mistake, i.e., fix my noema.

The key to fully grasping Husserl's view of demonstratives is held within his account of perception. West coast Husserlians, David W. Smith and Ronald McIntyre, both former students of Føllesdal, refer to Husserl's perceptual noema as the "acquainting sense", which is essentially the sense of a demonstrative (Smith and McIntyre 1982, p. 364). The acquainting sense is composed of two parts: (1) a generic component, the general way that an object is given in a perceived acquaintance, and (2) a specific component, which is the representation of the object itself on a particular occasion. The former is to be regarded as a sort of generic mode of presentation, e.g., the general ability to perceive in addition to the contents of mental demonstratives. The latter *specific* component is what we are most concerned with because it is the true sense of demonstratives; Smith equates this to what Husserl calls the "determinable X" (Smith 1982, pp. 181–184).

While Smith's distinction of the generic and specific components is helpful, the role of the X in perceptual acquaintance is in need of clarification. The X is the object "*simpliciter"* and is "in

abstraction from all predicates" (Husserl 1976 §131). It is specific, like Smith says, insofar as the object of a demonstrative reference changes relative to context. It is true that the content of the X varies across utterances; however, it seems to me that the X has an invariable aspect as well. Despite changes in experience, the X nevertheless focuses the various modes of presentation as belonging to the intentional object. This invariable aspect seems to correlate with the linguistic character of the demonstrative or its functional role.

This brief description of the X should raise a red flag. What exactly *is* the X? Is it the physical object, a percept, or is it a concept? And if the content of a demonstrative is variable then it should follow that its content is *not* purely cognitive; but rather, it is dependent on perceptual experience or other phenomenologies, definitely contrary to Pitt's view. The goal for the remainder of the paper is to argue that the X is a concept, *not* a physical object or a percept, and that if we pick apart the acquainting sense carefully, we shall see that Husserl's account of the content of demonstratives is compatible with Pitt's. I believe that making this comparison elucidates Husserl's explanation of demonstratives.

The acquainting sense as a whole helps to fulfill perceptual experiences, thereby increasing the degree of understanding. The role of the X is to focus attention on the object presented and determine the intentional relation to the object. It is *invariable* insofar as it is the linguistic character of the expressed demonstrative. Yet, it is *variable* with respect to the worldly context that helps to assign a referent. The predicates of experience change, but the X is what allows the subject to keep track of these changes by assigning them a unity or identity. Husserl writes in *Ideas* §131 regarding the X, "It must be distinguished from these (the predicates perceived), although it should not be set alongside them and should not be separated from them, as inversely they themselves are *its* predicates: inconceivable without it and yet distinguishable from it" (Husserl 1976, §131). The X is a function that allows the viewer to keep track of the presented object as it undergoes change or transformation.

A cause of confusion between Husserl and Pitt is that the noema and the acquainting sense include more than raw perception, namely, cognition as well. To unpack Husserl's view, let us examine the role of cognitive phenomenology in the noema and acquainting sense. Although Husserl does not explicitly mention cognitive phenomenology by name, he would definitely agree that each phenomenal type is distinct and separate from one another—seeing is different than tasting, which is different than hearing, etc. But he would maintain that we rely on multiple senses to gain knowledge from our experiences. Husserl seems to regard cognition as a higher order activity that trumps other types of phenomenological activities; a variety of phenomenologies may be constituents of the same total state of consciousness, but the non-cognitive phenomenologies cannot find their way into pure thought (McIntyre 2009, Interview):

> The phenomenologist can at will, through an act of reflection, change the intentional correlate of his act into an object of a second-order act. He can think of the thought rather than the object he is thinking about, and he then becomes aware that the thought was present all along (Dreyfus 1982, p. 101).

It seems to be the case that when we directly perceive an object, we can simultaneously have thoughts about it. Husserl approaches perceptual experience on a macro-level, that of a phenomenologically loaded experience, and so he describes the general way in which we interact with the world; this involves thinking and perceiving at the same time:

> But *separated* acts, as, for instance, two perceptions or a perception and a memory, can likewise close together in a 'harmonious' unity, and by means of the unique nature of this closing together, […] the Something […] is now consciously grasped as the same Something, or as in common accord the same 'object' (Husserl 1982, §131).

Although Husserl is not as precise and explicit as Pitt in describing

the unique role of cognitive phenomenology, he does not say that the noema blends together different types of phenomenologies; rather, he would view cognitive phenomenology as working side by side with the other types of phenomenologies.

Some examples will help to demonstrate this; let us revisit the unidentifiable aircraft scenario from earlier, only this time with the use of demonstratives and more terminology. Once more, unbeknownst to me, an aircraft will soon fly overheard and land nearby. While the sky is still clear, I think ^{T}That is going to land^{T} about nothing in particular. With a clear sky, the ^{T}that^{T} would not refer to anything and my noema would be empty. Surely I understand what was thought, the context is just too slight to help me determine a referent. As I see an object flying high in the sky I think the same thought as before, ^{T}That is going to land^{T}. My noema becomes more fulfilled with the addition of the object in the sky and I prescribe the object as being an airplane. However, since the object is so far away, I cannot be entirely certain of my identification. The mental content of ^{T}that^{T} is still the same as when the sky was clear, regardless of the extra sensory experiences in the context. The X focuses my attention on the same object of thought, namely, the thing that I expect to land.

When the aircraft descends and prepares to land, I realize that I was mistaken and that the aircraft is in fact a helicopter. My noema "explodes" as I realize my assumption was incorrect and my horizon of expectations change. Since the aircraft is closer to me, my perceptions are clearer, more veridical, and thus more fulfilling. The X component of the acquainting sense fixed my attention to the aircraft and I thought the same thought type throughout the episode. However, the change of context is responsible for ^{T}that^{T} to first fail to refer to anything, then refer to an airplane, and finally, a helicopter. Yet, ^{T}that^{T} still referred to the same aircraft while I was confused about the aircraft's type. The perceptual changes are what make the X variable, but it is an invariable function that helps to assign reference.

To regard the X as being *variable* with respect to a context and *invariable* with respect to its functional role should not be

a problem. The content of a demonstrative is still its linguistic meaning and it determines reference relative to context, which is inherently variable. The three examples above involve the same thought type, but each example has a different noema. Consider the following passage from Husserl pertaining to perceptual acquaintance and the X:

> We say that in the continuous or synthetic process of consciousness we are persistently aware of the intentional object, but that in this experience the object is ever "presenting itself differently"; it may be "*the same"*, only given with other predicates, with another determining content; "it" may display itself only in different aspects whereby the predicates left indeterminate have become more closely determined; or "the" object may have remained unchanged throughout this stretch of givenness, but now "it", the selfsame, changes, and through this change becomes more beautiful or forfeits some of its utility-value, and so forth (Husserl 1976, §131).[1]

In the cognitive sense, I thought different tokens of the same thought type throughout the episode. The qualia pertaining to my thought ^{T}That will land^{T} was the same across the three utterances, even though contextual changes altered my perceptual experiences. Thus, Husserl considers the X invariable as a function, yet variable when it forms a unity with context and determines a referent.

Until now, I have presented Pitt and Husserl as holding similar accounts of demonstratives. In order to test this claim, let us examine how Husserl would respond to the three objections that Pitt raises in opposition to his view. Those who maintain that demonstratives are referent dependent may assert the following: (1) two people cannot think the same ^{T}I^{T}-thought because they refer to different people, (2) in order to evaluate a particular demonstrative thought at another possible world, its content must be fixed as including its actual referent, and (3) if one does not know the referent of the demonstratives in a demonstrative thought then one

does not understand what was thought.

Again, Pitt says that if two people think ^{T}I'm hungry^{T}, then they are each thinking a token of the same thought type. The ^{T}I^{T} involved has its linguistic character that determines reference relative to each person. Demonstratives are not referent-dependent and each thought type has a unique phenomenology. We must recognize that the noema describes the partnership between pure thought and context. If we reflect on our experience, or more specifically, reflect on the acquainting sense, then we are aware that pure thought is present all along. The content of ^{T}I^{T} means something like 'the self that is being attended to'. When given a sufficient context, the ^{T}I^{T} now refers to the individual who thought it. Thus, in pure thought we can say that the two people are thinking tokens of the same thought type.

The case of modality raises the concern over whether or not what I've thought in the actual world is true or not at that world. For example, although I may truly think ^{T}I'm hungry^{T}, I might not be hungry at another possible world. Some say that the content of a demonstrative is rigid and includes its actual referent. This would require bringing along the individual for evaluations at possible worlds and therefore implies that demonstratives are not phenomenologically constituted. This is tricky for Husserl to address because once again, he must abstract cognitive phenomenology from the noema. Smith and McIntyre say:

> Now, a perception is always a definite intention, by virtue of its acquainting sense. And, on our account, it is rigid. Hence it is represented by a certain function that assigns the same individual to each world. But it does not follow that the perception must be individuatively definite. An individuatively definite intention of the same individual would be represented by the same function, but it may have an individuative sense that is not present in the perception (Smith and McIntyre 1982, p. 369).

To explicate, a perception is rigid because the acquainting sense is presented with a rich context, thus immediately allowing the X

to assign a reference. It is entirely possible to mistakenly identify an object, but mistakes concern the way in which we perceive predicates, not the identity of the object. The X function in a perceptual experience will assign these predicates an identity and track their changes, regardless of their correctness. Husserl would say that my thought, ⌜I'm hungry⌝, is rigid, but only in the sense that I could not have been someone else. On the other hand, a demonstrative thought with an empty sense would not individuate anyone or anything in particular because pure thought abstracted from the acquainting sense is left with just its linguistic character (ibid., p. 369). Husserl indicates that the noema is the unification of two concepts, the *"object simpliciter*" and the "*object as modally determined*". The modal form is part of the noema and is that which an act prescribes in a particular context (Husserl 1976, §131). Thus demonstratives are not referent dependent for Husserl, since modal evaluations of demonstratives concern referents and truth conditions, both of which require a context to refer.

In the third case of thinking ⌜That's a strange thing⌝ without referring to anything in particular, the X is still performing its role of directing one's thoughts towards a referent; only in this case, the context is too empty to determine one. Even though ⌜that⌝ is empty or thin, it still has content, namely, the linguistic role of selecting something that I am attending to. Smith and McIntyre seem to be acknowledging that the phenomenology of thought content is purely cognitive when they discuss the X outside of context, "…the X in a perceptual Sinn embodies the structure of the presentation of an object "itself". But an X is quite mysterious if it appears all by itself. There must be other items of Sinn that embody the way that object is ultimately presented and so are responsible for the presence of the X" (Smith and McIntyre 1982, p. 364). Outside of context, the X is left with its invariable aspect, i.e., its functional role in determining reference. Without a context, demonstratives will fail to refer to anything, however, if the acquainting sense is empty, it is not meaningless. The words themselves have meaning and can be understood. The subject understands the functional role or character of the demonstrative,

but is deprived of a context to successfully determine a referent.

A final problem that is still lurking in the background is what Pitt and Husserl deem to be the bearer of meaning, and what meaning is in general. Pitt says that the content of thoughts, including demonstrative thoughts, is purely cognitive, and that we gain understanding in degrees relative to context. On this interpretation of Husserl, the phenomenology of thought content is also purely cognitive, and like Pitt, the meaning of a demonstrative in thought is just its linguistic character. The noema or acquainting sense is an abstract entity, a function from mind to context, through which the referent is determined. Husserl considers the noema the bearer of meaning; however, the meaning or content of *pure thought* is different than the meaning or understanding that we gain through *experience* (Husserl 1976, §§131–132). Husserl notices that thoughts pervade the total conscious experience, and that cognition is separate and trumps other forms of phenomenological experiences (Dreyfus 1982, p. 101). Therefore, when Husserl says that the noema bears the meaning of an act and prescribes and object for the perceiver, he is entertaining two kinds of meaning at once, i.e., both pure thought content, namely, linguistic character, and higher degrees of understanding gained through experience.

Despite having different approaches to explaining demonstratives, it is evident that Pitt's view and the "west coast" interpretation of Husserl are compatible with each another. Since Husserl's stance on the subject of demonstrative reference is not obvious, I find it helpful to interpret his theory of perceptual acquaintance using Pitt's claims concerning the phenomenology of cognition. On this account, the X is a concept, insofar as its function is to assign and track identity. I believe that this reevaluation clarifies how the X in perceptual acquaintance is *invariable* with respect to its functional role—in this case, its demonstrative or linguistic character—and *variable* with respect to context.[7]

Notes

1. Demonstratives are terms whose referents change relative to context and usually require some sort of demonstration, gesture, or vocal emphasis to secure a referent, e.g., this, that, he, she. They are part of the broader family of terms called indexicals, whose referents also change reative to context, yet indexicals do not require an additional demonstration, e.g., I, today, yesterday, here, now. The problem with demonstratives is that they appear to require more than just the linguistic meaning and context to determine reference—they seem to require an additional demonstration.
2. Pitt uses the superscripted T-quotes to represent thoughts in the head; they function in the same way that quotes do for spoken utterances. I use his notation throughout the paper.
3. I use 'act' or 'mental act' interchangeably as meaning any sort of intentional activity, e.g., 'thinking about', 'wishing that', 'judging that', etc.
4. Interpreting the noema as a wholly abstract and conceptual is the salient feature of the "west coast" or Føllesdalian interpretation.
5. On the "west coast" interpretation, the noema and object distinction is akin to Frege's sense and reference.
6. The noemata are tokens of the same type. They have the same meaning, but are numerically distinct.
7. I'd like to thank Ron McIntyre for giving me clarification on his interpretation of Husserl, and David Pitt for instructing me on his view.

Bibliography

Dreyfus, Hubert L. (1982) "Husserl's Perception of Noema," in: Hubert L. Dreyfus (Ed), *Husserl, Intentionality, and Cognitive Science*, pp. 97–123 (Cambridge, Mass: MIT Press)

Husserl, Edmund. (1976) *Ideas General Introduction to Pure Phenomenology* (New York: Prometheus Books)

McIntyre, Ronald. Personal Interview. 6 Mar, 2009.

Pitt, David. (July 2004) "The Phenomenology of Cognition Or *What Is It Like to Think That P?*," *Philosophy and Phenomenological Research*, pp. 1–35

_____. "Demonstrative Thoughts" Unpublished, pp. 1–19

Smith, David W. (May 1982) "What's the Meaning of This?" *Nous* Vol. 16, No. 2, pp. 181–208

Smith, David W. and McIntyre, Ronald. (1982) *Husserl and Intentionality a Study of Mind, Meaning, and Language* (Kluwer Boston Inc.)

Cognitive Phenomenology and Demonstrative Thoughts

Michael Anderson

Abstract: Many have thought that demonstratives or indexicals are directly referring. That is they are a kind of linguistic finger pointing; they stand for the object they refer to. If this is right, any proposition containing a demonstrative or indexical is going to be (at least partially) individuated by the object it refers to. If thoughts are propositions, then any thoughts containing demonstratives or indexicals (henceforth, demonstrative thoughts) are also individuated by the objects they refer to. This thesis, that demonstrative thoughts are individuated by the objects to which they refer, I will call OD (short for the Object Dependence Thesis). OD is problematic for the theory that thoughts are phenomenally individuated because phenomenal qualities are mental properties, and presumably mental properties can only exist in the mind. However, if OD is true and a demonstrative thought has as its referent a non-mental object (concrete or abstract), that object cannot exist in the thinker's mind. Thus, if OD is true, demonstrative thoughts cannot be phenomenally individuated. David Pitt is one such philosopher who endorses the view that thoughts are phenomenally individuated. In his paper, *Demonstrative Thoughts*, Pitt considers three intuitions in support of OD and seeks to (at the very least) show that the intuitions are not conclusive as to the truth of OD. I will evaluate Pitt's treatment of the three intuitions. In addition, a consequence of Pitt's treatment is that two demonstrative thoughts that are the same, except that they have different referents, are still the same thought. This goes against my intuitions for many cases. I believe that there are cases in which difference of reference causes difference of thought. So I will offer a slightly different treatment of demonstrative thoughts that, I believe, is consistent

with the view that thoughts are phenomenally individuated but is also consistent with my intuition that there are cases in which difference of reference causes difference of thought.

1. Background

Within analytic philosophy, there is a traditional separation between thoughts and experiences. Thoughts are about things, that is, they are intentional. We might ask someone what they are thinking about, or more precisely, what is their thought about? And if their thought[1] is $^{T}2+2=4^{T}$, they would answer, "Well, my thought is about two plus two equaling four." On the other side, there is something it's like to have a phenomenal experience. If a person is having an experience of pain, we might ask them, "What does your pain feel like?" They may answer, "It hurts," and give more information by saying, "I feel this burning in my stomach," and if we have ever experienced what burning is like in or out of our stomach, we get some kind of idea of what their pain is like.

Those who hold to the traditional separation of thoughts and experiences further add that there is little to no overlap between phenomenal experiences and intentional content. They argue that thoughts have no phenomenal content, and phenomenal experiences have no intentional content. To ask, as the story goes, what a phenomenal experience is about is an inappropriate question. What is your pain experience about? Nothing; there is no "aboutness" for phenomenal experiences. On the other side, intentional thoughts have no "what-it's-likeness." What is it like to have your thought $^{T}2+2=4^{T}$? Well, nothing. Thoughts are about things; there is nothing it's like to have them, i.e. there are no phenomenal qualities of intentional thoughts.

This traditional view is assumed and taken for granted by many philosophers. There are many works in the philosophy of mind and language where this paradigm is assumed. However, in recent years, this doctrine has been challenged by philosophers within analytic philosophy; one of whom is David Pitt.

Pitt argues that this distinction is false, there *is* phenome-

nology within thoughts and there *is* intentionality about experiences. In his paper, *The Phenomenology of Cognition,* Pitt argues that thoughts are phenomenally constituted, and this phenomenal constitution makes up their content and individuates them. His argument is as follows: When I think $^{T}2+2=4^{T}$, I know what I am thinking. How do I know that? How do I know that I am not thinking ^{T}grass is green^{T}? Obviously $^{T}2+2=4^{T}$ and ^{T}grass is green^{T} are about different things, but how do I know when I am thinking one and not the other? The answer, says Pitt, is that they have phenomenal qualities. What it's like to think ^{T}grass is green^{T} is different than what it is like to think $^{T}2+2=4^{T}$, and I know what I am thinking by the phenomenal experiences of my thoughts, that is, I can identify them by what it's like to have them, and I distinguish them from other thoughts in the same way. I can do this because the phenomenal content is what constitutes the thought and makes it the thought that it is, which is distinct from the phenomenal content of other thoughts.

Even more, I know when I am having intentional, phenomenal thoughts; I am able to distinguish such phenomenology from other types of phenomenology (visual, auditory, tactile, olfactory, emotional, etc.) because there is something it is like to have this intentional phenomenology that is different than what it is like to have the other types. Analogously, I am able to distinguish auditory phenomenology from visual phenomenology because there is something it is like to have an auditory experience that is different than what it is like to have a visual experience. Such ability suggests that auditory phenomenology is unique and different than visual phenomenology. In the same way the ability to distinguish intentional phenomenology from the other types suggests that intentional phenomenology is unique and different from the others. We will call this distinct phenomenology cognitive phenomenology.

Pitt puts forth the following thesis:

(P) "Each type of conscious thought—each state of consciously thinking that p, for all thinkable

> contents p—has a proprietary, distinctive, individuate phenomenology."[2]

By "proprietary," we mean that cognitive phenomenology is different from any other phenomenal state. That is, what it is like to think ^{T}that p^{T} is different than what it is like to have any other experience. Each type of conscious thought has a "distinctive" phenomenal content, what it is like to think ^{T}that p^{T} is different than what it is like to think any other thought. The phenomenology of a thought is what constitutes a thought's representational content, making it the thought it is; it is "individuative."

This view that thoughts are essentially composed of their individual, cognitive, phenomenal content is naturally going to clash with other theories. One such theory is Direct Reference Theory. DRT holds that proper names, pure indexicals ('I', 'today', 'tomorrow', 'actual', 'present', and (perhaps) 'here' and 'now'), true demonstratives('he', 'she', 'his', 'her', and 'that'), and (perhaps) natural kinds are directly referential. They do not stand for any description, cluster of descriptions, mode of presentation, concept, etc.; they just stand for the actual object (referent); that is, they directly refer to the object. The statement "George Clooney is an actor" just means <the Guy$_{\text{George Clooney}}$, the property$_{\text{'is an actor'}}$>. If one were to ask, "What do you mean by 'George Clooney?'" I might respond by giving some description of him (say, the guy in Oceans 11) or by showing a picture of him. But those are merely ways of picking out the referent. The meaning of 'George Clooney' is not the description: 'the guy who was in Oceans Eleven'. Rather, I am using the description as a means of pointing out the guy, the actual person/object. The "meaning" of proper names is just their referents.

In the case of demonstratives, folks like David Kaplan hold that they are also directly referential. When I say "I am tired," the contribution that 'I' gives to the meaning of the statement is, like proper names, just the object. So 'I am tired' just means <the guy$_{\text{me}}$, the property$_{\text{is tired}}$>. For the expression 'that is red', 'that' refers to the object that I am picking out through some sort of

method (like pointing with my finger). Like 'I', the only thing 'that' contributes to the meaning is just the object. Parallel treatments can be given to the other common demonstratives. Unlike proper names, however, says Kaplan, demonstratives do have something like senses; Kaplan calls these characters. For example, the character of 'I' is 'the relevant speaker/thinker/writer," and the character of 'that' is "the relevant object," but these characters do not contribute anything to the meaning of a given statement; it is the object picked out by the demonstrative that contributes meaning.

Another way to talk about this is to say that proper names, demonstratives, indexicals, and (some believe) general terms are what Saul Kripke calls rigid designators. A rigid designator is a term that refers to the same object in every possible world in which that object exists and refers to nothing in worlds where that object does not exist. Take the following two sentences:

1. The lead actor in Oceans Eleven could have been Ben Affleck.

2. George Clooney could have been Ben Affleck.

1 is clearly true; it is possible that Ben Affleck could have had the lead role in Oceans Eleven. George Clooney has that role in the actual world, but there is a possible world in which Ben Affleck has that role. What about 2? It seems equally clear that it is false. There are no possible worlds in which George Clooney exists where he is Ben Affleck. What's the difference? Names are rigid designators while descriptions are not. When we take the description 'lead actor in Oceans Eleven' to another possible world, all that is required of the relevant object is that it satisfies that description. So any object, with the capacity to do that, can be the relevant object. When we take 'George Clooney' to another possible world, there is only one object that can be depicted by 'George Clooney', and that is the guy, George Clooney himself; likewise for 'Ben Affleck'. In most cases, we are not asking whether George Clooney could have been named Ben Affleck. We

are asking whether the guy George Clooney could have been the guy Ben Affleck? Clearly not.

Similarly for demonstratives and indexicals, once we attach an indexical to an object it becomes rigid. Consider the next two sentences:

1. The person sitting in front of this computer is Barack Obama.

2. I am Barack Obama.

Both 1 & 2 are false. But 1 could have been true; whereas 2 could not. There are possible worlds in which Barack Obama satisfies the description 'sitting in front of this computer'. That is, there are possible worlds in which Barack Obama is sitting in front of this computer. But there are no possible worlds in which I am Barack Obama.

It does not take much consideration to see why such a theory would be problematic for Pitt. If thoughts are phenomenally constituted, then there can be no objects in the meaning of ^{T}I am hungry^{T} because there can be no phenomenality outside of the mind, and there can be no objects inside the mind. In other words, DRT about proper names and demonstratives means that the thoughts ^{T}I am hungry^{T} or ^{T}John is smart^{T} cannot be phenomenally constituted.

2. Pitt's Treatment

For our purposes, we are concerned with Pitt's treatment of demonstrative thoughts. His main purpose in this paper is: (1) to consider three intuitive reasons for OD (2) to "blunt these intuitions" and (3) "to defend a conception of demonstrative thought individuation that is entirely phenomenal and internalist."[3]

Let us consider each of the three intuitive reasons as well as Pitt's attempt at "blunting" them.

(1) **The "What is thought" motivation.** It is intuitive to think that when Fred thinks ^{T}I am hungry^{T} and when Karl thinks ^{T}I am hungry^{T}, they are thinking different thoughts because Fred's

thought is about Fred and Karl's thought is about Karl. Likewise, when I think ^{T}This is a nice place^{T} while being under a tree and again while in a house, it is intuitive to think those thoughts are different because the first is about the place under the tree, and the second is about the place in the house. This belief, that they are two different thoughts, is further reinforced by the fact that they can have two different truth values. When Fred thinks ^{T}I am hungry^{T}, it could be true, while when Karl thinks it, it could be false, and visa versa. It's odd to think that they could be the same thought and have different truth values. This suggests that difference of reference entails difference in the content of thoughts.

Pitt rejects this reasoning that difference in reference and possible difference in truth value entails difference in thought. He gives us some examples. Suppose I think in 1972 ^{T}The president of the United States is a fool^{T} and then I think this again in 2002. Despite the referents being different and possible difference in truth value, Pitt's intuition is that the thought is the same. Likewise, suppose in 1972 I think ^{T}Dogs are quadrupeds^{T} and again in 2002, during that time new dogs and quadrupeds have come in and out of existence, so the extension (referent) of both is different. Again Pitt's intuition is that despite the difference of referent and possible difference of truth value, the thoughts are the same.

You might object to Pitt's reasoning and think that the content of thoughts contain time indicators. So when I think ^{T}Dogs are quadrupeds^{T} in 1972, what I really mean is ^{T}Dogs are quadrupeds in 1972^{T} and likewise for 2002. However, this (eternalism) is implausible. You can think ^{T}Dogs are quadrupeds^{T} without knowing or even thinking about what date it is. In addition, the beliefs that the contents include place indicators (ubiquitism) or world indicators (necessitarianism) are also implausible for similar reasons.

You might argue that difference of reference entails different thoughts by appealing to the sense-determines-reference principle. When Fred thinks ^{T}I am hungry^{T} and Karl thinks ^{T}I am hungry^{T}, they must have different senses (contents) because they have different referents. If they were the same content, they would

have the same referent because sense determines reference.

This is a misunderstanding of the description theorist's principle. It is not really that sense determines reference, it is that sense + context determines reference. Thus, in the ^{T}Dogs are quadrupeds^{T} case, "in order to assign an extension of ^{T}Dogs^{T}, one must specify a world or a time (context)."[4] For demonstratives, the context required to determine reference is even more specific. Demonstrative thoughts can change their extension relative to time, place, world, and even things like perceptual-attentive states of individuals. Thus difference of reference does not entail difference of sense (content).

(2) **The modal motivation.** Suppose I think ^{T}I am hungry^{T}. Assuming this is true, we might ask of my thought, "Could it have been false?" or, to put it differently, "Is there a possible world in which it is false?" There are two ways in which it could have been false. First, 'I' could have referred to Fred, and he had already eaten. Second, it could have been false if I had eaten lunch, and therefore was full. To put it differently, the thought could be wrong if (1) there is a possible world where 'I' refers to Fred, and he is not hungry, or (2) there is a possible world where I ate lunch and am not hungry. Normally when we ask "Could it have been false?" we are not asking a question in which (1) would be a satisfying answer. We do not care about possible worlds in which someone else is not hungry? We care about whether or not there is a possible world in which I (me) am not hungry? What's relevant to the question is the referent of the thought.

Pitt believes it is inappropriate to ask if a thought could have been true in a possible world; but rather, we should be asking, "Could it have been true *at context C* in a possible world?" The thought ^{T}I am hungry^{T} can only be true or false in this, or any other world, relative to context. Consider the thought ^{T}The mayor is in his office^{T} when we ask "Is that thought true?" we cannot give an answer until we have specified a context. If we think the thought in Los Angeles, the referent is Antonio Villaraigosa. If we think the thought in New York, the referent is Michael Bloomberg.

When I ask whether the thought ^{T}I am hungry^{T} is true in

possible world w? "The answer depends upon which context in w is selected … I can ask if what I thought would be true of me or of you or of anyone else at another possible world w."[5] Further, I can ask whether it is true of different times, places, etc. depending on the context I choose. That depends on which person, place, time, etc. I am interested in.

So far we have, from Pitt, that demonstrative contents do not have their referents as part of the meaning, nor are they rigid designators, apart from context. So neither the *what is thought* nor the *modal* motivations "militate against the thesis that demonstrative thoughts are entirely phenomenally individuated."[6]

(3) **The demonstrative without referent motivation.** It seems intuitive that unless you have the referent picked out, you cannot understand the demonstrative thought. You can't understand 'this is ridiculous' unless you understand what 'this' is. This is also true of thoughts. The thought ^{T}this is ridiculous^{T} seems incomplete until you have specified what ^{T}this^{T} is.

Pitt responds that there is a difference between a thin sense (content) and an incomplete sense. An incomplete sense is an expression like 'is red'; it is as if the sense has a hole in it. If demonstrative thoughts are incomplete senses, then it would be true that we cannot understand the thought unless we specify the referent. However, this is not how Pitt thinks we should treat of demonstratives without determined referents. Rather, we should think of them as thin but complete. It is true that 'that' without a referent is not very informative; but it does have meaning. Consider this demonstrative expression: 'that large, round, green, ball, in the corner', as we erase each term, the content becomes less and less informative, but it does not follow from this that the thought becomes less and less understandable. Even when we arrive at 'that', it still has a fully complete meaning; it is just thin (or minimally informative).

3. Evaluation

In what follows, I will present a general response. This response is

not an in-depth treatment of Pitt's paper, but rather some general observations. I argue that those who are firmly entrenched in the OD camp, so to speak, are not going to find Pitt's treatment very convincing. I will call our ODist, Kenny. The responses that I have are those that I think Kenny will have.

There are two ways to understand Pitt's project. One, he is confronting the intuitions believed to be supportive of OD by suggesting other explanations for those intuitions. Two, he is offering counter-intuitions that would suggest that demonstrative thoughts are not referent involving. For both projects, Pitt relies heavily on examples and suggests that his treatment of those examples is analogous to how we ought to treat demonstrative thoughts. If the above characterizations of Pitt's strategy are correct, I argue that a lot hinges on: 1) the intuitiveness of his treatments of those examples and 2) the transferability of their intuitiveness to demonstrative cases. Those examples include: ^{T}The President of the United States is a fool^{T}, ^{T}The mayor is in his office^{T}, and ^{T}dogs are quadrupeds^{T}.

Pitt says of the two thoughts (in 1972 and 2002) ^{T}The President of the United States is a fool^{T}, "I have a strong intuition that I have thought precisely the same thing on both occasions, in spite of the fact that in 1972 I'm referring to a different fool than the one I'm referring to in 2002,"[7] and implicitly he has the same intuitions of the two thoughts ^{T}The mayor is in his office^{T} (in L.A. and in New York). I argue that both thoughts can be interpreted in two different ways. Under the first interpretation, Kenny will share Pitt's intuitions, but the intuitiveness in this interpretation does not obviously transfer to demonstratives. The second interpretation is transferable, but Kenny will no longer share Pitt's intuitions.

Those two interpretations are as follows:

'The President of the United States is a fool'

1. The guy who holds the office of president, whoever that may be, is a fool.

2. George W. Bush (or Richard Nixon) is a fool.

'The mayor is in his office'.

1. The guy who holds the office of mayor, whoever that may be, is in his office.

2. Antonio Villaraigosa (or Michael Bloomberg) is in his office.

If it is 1, then Kenny will agree with Pitt that the content of my thought tokens (in 2002 and 1976 or in L.A. and in New York) is the same. However, this does not motivate him to accept that when I think ^{T}I am frustrated^{T} and when you think ^{T}I am frustrated^{T}, we are thinking the same thought, because the terms 'the president of the United States' and 'the mayor', when understood in this sense, are obviously not rigid. This is not so obvious for the expression 'I'. If it is 2, then the terms 'the president of the United States' and 'the mayor' are rigid. However, Kenny will no longer agree that the thoughts (in 1972 and 2002 or in L.A. and New York) are the same thoughts.

For the ^{T}dogs are quadrupeds^{T} thought, we are on shaky ground. Both 'dog' and 'quadruped' are general terms. General terms are often considered rigid. But exactly how they are rigid designators is unclear. What does seem clear is that general terms, if rigid, are rigid in a significantly different way than demonstratives are (allegedly).

Scott Soames suggests that a general term is rigid if it expresses an essential property. That is, a general term is rigid when it expresses a property that any object which possesses it cannot fail to have.[8] For example, the general term 'human being' expresses a property(s) that if anything is a human being it must have that property(s). There is no possible world in which a human being could be a rock because rocks cannot have the property(s) essential to being human. There are, however, possible worlds in which a human being could be a doctor, because doctors can have the property(s) essential to being human. Ultimately, however, Soames is skeptical about this view because there are problems when one tries to formalize identity statements containing

predicates in the same way as one would for singular terms. He concludes that there is no extension of rigid designation from singular terms to general terms that captures what is needed for the rigidity of general terms. "In order to avoid confusion, it may be advisable to reserve the terminology of rigidity exclusively for singular terms."[9]

Nathan Salmon rejects Soames' idea that a general term's being rigid means that it is an expression of an essential property, because it eliminates colors as being rigid. For example, no object that is red is red essentially. This goes against Kripke who labeled 'red' as rigid. He also rejects it because it "does not provide any obvious candidate to be the rigid designatum of a predicate like 'is a tiger.'"[10] He also disagrees with Soames that there is no natural extension from the rigidity of singular terms to the rigidity of general terms. His candidates, to be the rigid designatum, are universals. Just as proper terms designate individuals, general terms designate universals.

Salmon's theory seems to be more close to the spirit of what Kripke was going for (as Soames probably would agree). But I wonder if universals aren't going to ultimately amount to the same thing as essential properties. However, my main reason for expositing both philosophers is to point out that both agree in at least one thing: if general terms are rigid, they are rigid in a different way than 'I', 'he', 'this', 'that', etc. are. "But insofar as the extension of the general term 'tiger' is the class of actual tigers, it is clear that the term does not rigidly designate its extension, since the class of tigers in one possible world may differ from the class of tigers in another world."[11] Whereas the traditional view of demonstratives is that they do rigidly designate their extension (referent).

All this is to say that there are significant disanalogies between Pitt's examples and demonstratives. These disanalogies will likely make if difficult for Kenny to take the leap from Pitt's examples to demonstratives.

4. An Alternative Version

So it seems that Kenny, and his kind, would find Pitt's paper unconvincing. However, something tells me Pitt is not overly concerned by this. I think those who would find Pitt's paper most convincing are those who are not wholly committed to the doctrine of direct reference; and perhaps those who are considering cognitive phenomenology but are in need of a coherent theory of demonstratives that does not commit one to OD. In regard to this, Pitt is successful.

Let us assume this is true, and Pitt has done an adequate job of "blunting" the intuitions that suggest demonstrative thoughts are referent involving (given the audience he is writing to). What has he accomplished? Well, he has removed one reason for believing that thoughts could not be phenomenally constituted. Further, he has suggested that Fred and Karl's thought ^{T}I am hungry^{T} are the same thought even though they have different referents. Suppose we are prepared to accept cognitive phenomenology that thoughts are phenomenally constituted and do not contain the objects they refer. Are there any reasons, apart from the three arguments, to think that Karl and Fred's thoughts are different? [12] Or to put it a bit differently, is there any room in cognitive phenomenology to suggest that they are different thoughts?

I think there is reason to believe that Karl and Fred's thoughts *are* different, apart from the three arguments. In fact, I think one can appeal to Pitt's own theory of individuation to suggest that Fred and Karl's thought are different. I will present an alternative to Pitt's theory of cognitive phenomenology that can accommodate for the difference between certain demonstrative thoughts. Further, it is my aim to accomplish this while preserving as much of Pitt's original theory as possible.

According to Pitt's theory, I know when I am thinking ^{T}2+2=4^{T} because there is something it is like to think ^{T}2+2=4^{T} that is different than what it is like to think, say, ^{T}dogs are quadrupeds^{T}. That is, they are phenomenally different. If thoughts are phenomenally constituted, then it follows that the difference in phenomenal

content between $^{T}2+2=4^{T}$ and ^{T}dogs are quadrupeds^{T} entails that they are different thoughts. What about Fred and Karl's thought ^{T}I am hungry^{T}? I suggest that they are different for precisely the same reason. There is something it is like for Fred to think ^{T}I am hungry^{T} that is different than what it is like for Karl to think ^{T}I am hungry^{T}. And like the $^{T}2+2=4^{T}$ / ^{T}dogs are quadrupeds^{T} case, the difference in phenomenal content between Fred's thought ^{T}I am hungry^{T} and Karl's thought ^{T}I am hungry^{T} entails that they are different thoughts.

Pitt's response to my intuitions would be that what is really different is my other types of experience, but the content of the thought (the cognitive phenomenology) remains the same. I'm just not so sure that explains my intuitions. Further, I find it implausible that the other types of phenomenology do not affect the content of our thought.

Let me say that Pitt is not completely clear here. There are times when he says things like "Your token of the concept $^{T}I^{T}$ has exactly the same (phenomenal) content as mine, or anyone else's"[13] and also, "The content of the demonstrative concepts themselves, as well as the thoughts they are constituents of, will be *the same* on both occasions—just as the content of $^{T}I^{T}$ is the same on every occasion of its use."[14] These statements suggest that the other phenomenologies have no bearing on the content of demonstrative thoughts, because, surely, the other phenomenal qualities have changed, but according to Pitt, the content has remained the same. He says further that the other experiences should be included in the contextual factors that determine the reference.

However, there are other statements which seem to suggest that the content of a thought is changed when there is a difference in the other phenomenal qualities. In response to the third reason for believing that demonstrative thoughts are referent involving, Pitt gives the example of the thought ^{T}I don't like it here^{T} when the thinker is in a pitch dark room. He says, "If someone turns the light on, your understanding of 'here' would increase but only in the sense that *further* thoughts would be directed toward the

place where you are, and would be known to be about the same place."[15] Pitt seems to be suggesting that when the visual experience changes, the thought changes. Likewise, he says, "Someone who has no idea who Dick Cheney is could still think ^{T}The vice president is evil^{T}. If he knew who Dick Cheney is—what he looks like, where he's from, what he's done—that thought would entrain more information. But this further information would be in the form of *distinct thoughts.*"[16]

In any case, the most reasonable reading of Pitt's position is that he does not think a difference in the other sensations affects or makes a difference in the thought content. And this is where I disagree; I think that the other sensations have a causal relationship with the thought content.

David Woodruff Smith seems to have something similar in mind in his paper *What is the meaning of 'this'?*, as evidenced in the following quote: "The mode of presentation on which demonstrative reference is based, for the indicated use of 'this', is simply that of perception, or perceptual acquaintance. By uttering 'this' one refers to an object in one's current perceptual purview, and the structures of sense invoked by the utterance embody the mode of presentation of the object before the speaker's consciousness in perceptual acquaintance."[17] Joseph Levine also seems to be saying some of the same things in his paper, *Demonstrative Thought*. "There is a mediation, the mediation is that provided by perception itself. The idea is that mental demonstration selects its object through a kind of deferred ostension. The md (mental demonstrative) functions as a kind of pointer to the perceptual representation of the object, with the semantic value of whatever it is that the percept is a percept of."[18]

My theory differs from Smith's in the following important way. I am not suggesting that percepts *are* the sense or mental content of demonstrative thoughts. Neither am I suggesting, as Levine does, that the mental content of demonstrative thoughts is "a device for selecting the object to which the predicate applies."[19] I agree with Pitt that the mental content of demonstrative thoughts is the cognitive phenomenal qualities of those thoughts and only

the cognitive phenomenal qualities; or, as Pitt puts it the cognitive qualitative character of the thought content is proprietary. However, unlike Pitt (and perhaps more like Levine and Smith), I believe that cognitive phenomenology is heavily influenced by percepts which are captured by the other forms of phenomenology.

Note: I am not suggesting that the cognitive phenomenology is dependent on the referent as is argued when one appeals to the sense-determines-reference principle. In fact, I believe it is possible to have the same percepts from two distinct referents. That is, the thought about one referent is the same as the thought about the other, because their percepts are the same. The way in which we tell whether the thoughts are the same or different is due to cognitive distinguishability. Take the following scenarios:

(1) A. You are in a room and see a red book on a table and think ^{T}that is a book^{T} and then you look over to a shelf and see a blue book and you think ^{T}that is a book^{T}. In this case, the thoughts are cognitively distinguishable because the percepts are distinguishable. Thus, the thoughts are different.

B. You are in that same room and see a red book and think ^{T}that is a book^{T}, then you leave for a moment and while you are gone someone replaces the red book with a perceptually identical one. You think ^{T}that is a book^{T}. In this case, the thoughts are the same because the percepts are the same.

(2) A. You think ^{T}I am hungry^{T} and, at the same time, Karl thinks ^{T}I am hungry^{T}. I suggest that the thoughts are different because the percepts are different. If there was some weird quantum event that allowed you to be in your head and Karl's head at the same time when both you and Karl thought ^{T}I am hungry^{T}, those thoughts would have been cognitively distinguishable.

B. Imagine that you are on earth and there is a twin-you who is on Putnam's twin-earth. Imagine that twin-earth

is qualitatively identical to earth, and twin you is qualitatively identical to you. The only type of difference between twin-earth and earth and twin-you and you is quantitative. You and twin-you both think at the exact same time ^{T}I am hungry^{T}. Again had the weird quantum event happened which allowed you to be in your and "twin-you's" head at that moment when both of you thought "I am hungry^{T}, the thought would have been cognitively indistinguishable; and, so, the thoughts would be the same.

(3) A. You are under a tree and you think ^{T}this is a nice place^{T}. Then you go into a house and think ^{T}this is a nice place^{T}. The percepts would be different, so the thoughts would be cognitively distinguishable; and, therefore, different thoughts.

B. You are in a room and you think ^{T}this is a nice place^{T}. Then you fall asleep and, while you are asleep, your friends decide to play a trick on you by moving you to a different building where the room looks, smells, sounds, etc. exactly the way the other room did, and you think ^{T}this is a nice place^{T}. Again, the percepts are the same and so the thoughts are the same.

We might say that the percept for ^{T}that^{T} is perhaps the visual experience in the case of an object before you, or the auditory experience for perhaps ^{T}that barking dog^{T}, or an olfactory experience for ^{T}that smell^{T}, or a tactile experience for ^{T}that object^{T} if someone is blindfolded. Perhaps, even an emotional experience that one has in reference to ^{T}this place^{T}. It is important to note that the percepts which have a causal influence over the mental content of ^{T}I^{T} seems to be a bit different, or perhaps not as straight forward as the percepts for all the other demonstratives. We might say the thing which has a causal relationship with the cognitive content of ^{T}I^{T} thoughts is something like "self-awareness," which would seem to include an awareness of one's internal qualities and

one's relational qualities to other objects. It is plausible that such self-awareness comes from the collection of all (or most) types of phenomenal experience.

Also, I want to make it clear I do not believe that the only content cognitive phenomenology has is what is affected by the other forms of phenomenal experience. Pitt discusses a woman from college who had no sense of smell. He says that the terms denoting smells are not completely meaningless to her; I agree with him. In an analogous way, cognitive phenomenology still has content independent of the affection of other experiences. However, the content is thinner without the affections of percepts. Also, not all thoughts are equally affected. There are some thoughts (such as thoughts about truths of logic) that are very minimally affected (if at all) by percepts. On the other hand, lots of thoughts stipulating facts about the world are heavily affected.

Despite these differences, I still think that my variant is true to the spirit of Pitt's theory. According to my theory Pitt's thesis (P) is still intact; we just need to add clauses to it:

1. Proprietary: cognitive phenomenology is different from any other phenomenal state. + (Cognitive phenomenology is affected by the other phenomenologies.)

2. Distinctive: what it is like to think ^{T}that p^{T} is different than what it is like to think any other thought. + (Part of what contributes to this difference in cognitive phenomenology is due to the difference in other phenomenologies.)

3. Individuative: the phenomenology of a thought is what constitutes a thought's representational content. It makes it the thought it is. + (The phenomenology that constitutes the thought is affected by other phenomenologies.)

I find the view that thoughts are phenomenally constituted to be a very attractive one. And I have tried to show that this view is not necessarily at odds with the intuition that thoughts like ^{T}This is really beautiful^{T} when referring to the rose, is different than the thought when referring to the painting. I have argued that

when we allow the other kinds of phenomenologies to have influence over the content of our thoughts we can accommodate the intuition that they are indeed different thoughts. I conclude with one last example. Suppose a family is about to go on vacation to Hawaii and the mom (having been to Hawaii several times) thinks ^{T}Hawaii is so beautiful^{T}. So she says to her son, "You will love Hawaii; it is so beautiful!" Further, suppose her son has never been to Hawaii, nor has he seen any pictures. But later that night, the son thinks to himself ^{T}Hawaii is so beautiful^{T}. It doesn't seem unreasonable to view the mom's thought not merely as an enrichment of her son's, but in fact, a different thought.

Notes

1. Pitt introduces "thought quotes" which are superscripted, uppercase 'T's that work like normal quotation only they quote thoughts rather than verbal expressions. I will continue this procedure.
2. Pitt 2004, p.5
3. Pitt forthcoming p. 4
4. Pitt forthcoming pp. 8–9
5. Pitt forthcoming p. 12
6. Pitt forthcoming p. 13
7. Pitt forthcoming p. 6
8. Soames, 2002
9. Soames, 2002, p. 263
10. Solomon 2005
11. Salmon 2005 p.118
12. I am not suggesting that Pitt considered the three arguments to be exhaustive of all the arguments that would lead one to believe that they are different thoughts. I do believe, however, that he took the three to be typical of arguments appealing to principles based on either DRT or the sense determines reference principle. I am not asking if there are other arguments that appeal to those principles. But rather I am asking if there are any reasons for believing, that they are different thoughts not based on those principles.
13. Pitt forthcoming p. 10
14. Pitt forthcoming, P. 19
15. Pitt forthcoming, p. 14–15

16. Pitt forthcoming, p. 15
17. Smith (1982) p. 188
18. Joseph Levine forthcoming p. 15
19. Levine forthcoming p. 14

Bibliography

Kaplan, David (1989): "Demonstratives," in J. Almog, H. Wettstein, and J. Perry (eds.), *Themes From Kaplan*, pp. 481–563 (Oxford: Oxford University Press)

Kripke, Saul (1980): *Naming and Necessity* (Cambridge, MA: Harvard University Press)

Levine, Joseph (Forthcoming): "Demonstrative Thought."

Pitt, David (Forthcoming): "Demonstrative Thoughts."

Pitt, David (2004): "The Phenomenology of Cognition Or What Is It Like to Think That P?" *Philosophy and Phenomenological Research*, Vol. LXIX, No. 1

Smith, David Woodruff (1982): "What's the Meaning of 'This'?" Source: *Noûs*, Vol. 16, No. 2 pp. 181-208

Soames, Scott (2002): *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity* (Oxford: Oxford University Press)

Solomon, Nathan (2005): "Are General Terms Rigid?" *Linguistics and Philosophy*, Vol. 28, No. 1

Ideal and Nonideal Theory In Reflective Equilibrium

April Van Bibber

Introduction

"What is just?" This is the question behind a plethora of human decision-making about interaction between individuals, groups, and even nations. In the opening chapter of *A Theory of Justice,* John Rawls articulates the importance of justice, "justice is the first virtue of social institutions, as truth is of systems of thought" (Rawls 1999, p. 3). Trust withers where injustice thrives. Ultimately, where injustice becomes the rule rather than the exception, the social fabric is weakened, and the benefits resulting from relations of mutual trust necessarily decline. Recognition of the serious consequences that often result from unjust social and political conditions provides a powerful motivation for those concerned about human dignity to carefully consider the ways in which the principles of justice are derived and how closely institutions and policies match the ideals they express. To do this, of course, seems to require the articulation of ideals upon which, in a liberal democracy, an overlapping consensus of citizens might agree. While the answer to the question, "What is just?" usually begins with an intuition, philosophers have long recognized the value in articulating precise principles of justice so that they can be examined and improved, and, along the way, offer clear and concise policy objectives for theoretical reasoning about how to deal with injustice. This paper represents an inquiry into a methodology for specifying the ideals from which such principles are derived.

Agreeing with Rawls, John Simmons echoes the belief that "political philosophy should begin…with the defense of those

(moral) principles that ought to constrain the design and operation of the basic structure of each society" (Simmons 2010, p. 7). However, the decision to begin with ideal theory for the purpose of presenting nonideal theory with its task is not without its critics. A common criticism is that ideal theory fails to identify important injustices , such as oppression and relations of domination that, intuitively, ought not to exist in a just society. Moreover, Charles W. Mills' examination of biases in the perspective of ideal theorists leads him to conclude that ideal theory may be simply unnecessary because it, "can indeed be thought of as in part *ideological*, in the pejorative sense of a set of group ideas that reflect, and contribute to perpetuating, illicit group privilege" (Mills 2005, p. 166).

Unless one agrees that some injustice is acceptable in a realistically ideal society, which I am not prepared to do without reservation, principles of justice must contribute to the identification of injustices such that their eradication defines one of the objectives of nonideal theory. If principles of justice are unable to identify injustice, then they will fail on two counts. First, they will fail to describe an ideal standard upon which an overlapping consensus may be reached, especially those on the receiving end of any particular injustice left unidentified and unaddressed. Second, they will fail to contribute substantially to one of the objectives of nonideal theory: to theorize corrective policies for specific injustices. Although one might think that strict compliance to principles of justice ought to be effective in eliminating injustice, this is not necessarily the case. Lisa Schwartzman expresses a view similar to Mills' discussion of bias as she explores the claim, made by some, "that although rights can be used in arguments for women's equality, they can also function to uphold the power of privileged groups" (Schwartzman 2007, p.15). This is not a narrow, feminist claim, but one that identifies a significant breach within the mechanisms for distributing equality in a liberal democracy. If true, it represents critically important information that ought not to be overlooked in ideal theorizing. Thus, my thesis is that the proper methodology for stipulating principles of justice is not a one-way

flow of information from ideal to nonideal theory, but an ongoing exchange between the two.

Within the context of the above controversy, I will argue for a methodology that provides an arena, similar in structure to the space produced by the overlapping of two circles in a Venn diagram, in which ideal and nonideal theorists may work together, each contributing positively to the structure and content of the other. Such an arena provides for a lively and productive symbiotic relationship between parties mutually committed to promoting the continuing identification and eradication of political and social injustices. Furthermore, it is in just such an arena that the objections to ideal theory articulated by Mills and Schwartzman can be addressed.

In Section (1), I will summarize Rawls' theory of justice, clarify his distinction between ideal and nonideal theory, and introduce his notion of reflective equilibrium. In Section (2), I will explore Simmons and Mills on the question of the priority of ideal and nonideal theory, including the content and tasks of each. In Section (3), I will argue that characterizing the relationship between ideal and nonideal theory as symbiotic is more productive than prioritizing their tasks. Finally, in Section (4), I will argue that a methodology for a symbiotic relationship between ideal and nonideal theory is best articulated by Rawls' notion of reflective equilibrium.

Section 1

In *A Theory of Justice*, John Rawls rejects a utilitarian view of justice. The "greatest good for the greatest number" sacrifices too much for anyone who believes that "each person possesses an inviolability founded on justice that even the welfare of society as a whole cannot override… in a just society the liberties of equal citizenship are taken as settled; the rights secured by justice are not subject to political bargaining or to the calculus of social interests" (Rawls 1999, p. 3–4). He also rejects intuitionism, for the most part, and goes on to describe a new method for defending

principles of justice. His method, he claims, represents a justificatory improvement over the use of intuitions by enabling us to trace how we arrived at the principles of justice, and allowing for their modification when necessary. Rawls offers a two-pronged argument to support his principles of justice. In the first prong, the hypothetical contract argument, he claims that under conditions conducive to a fair hypothetical contract for principles of justice, the parties would choose his. In the second prong, the reflective equilibrium argument, he claims that the principles chosen are those that best fit our collective considered judgments of justice under ideal epistemic conditions in which we have access to whatever knowledge is necessary for making critical judgments pertinent to assessing principles of justice. Originally limited to "a closed system isolated from other societies," Rawls later offers an extension of his principles of justice to include the domain of international relations (Rawls 2002, p. 7). His purpose is to inform a realistic vision of a global utopia free of, "unjust war and oppression, religious persecution and the denial of liberty of conscience, starvation and poverty,…genocide and mass murder" (Rawls 2002, p.7).

Rawls' "realistic utopia" is the product of ideal theorizing about "the best social arrangement that we can realistically expect to achieve" (Rawls 2002, p.8). He offers a distinction between ideal and nonideal theory in order to limit his discussion to "principles of justice that would regulate a well-ordered society" (Rawls 2002, p.8). For Rawls, a well-ordered society is one in which the citizens obey the law and generally agree with and uphold its principles of justice. Thus, for Rawls, ideal theory attempts to describe a realistic utopia in which there is an assumption of strict compliance to principles of justice. Partial compliance (nonideal) theory, in Rawls' view, "studies the principles that govern how we are to deal with injustice" (Rawls 1999, p. 8). In ideal theory, then, strict compliance to principles of justice is the distinguishing characteristic of a utopia or well-ordered society. Narrowing the concept of a utopia to one that is *realistic,* Rawls claims, "extends what are ordinarily thought to be the limits of practical political possibility

and, in so doing, reconciles us to our political and social condition" (Rawls 2002, p. 11). He goes on to clarify his task, asking "What would a reasonably just constitutional democracy be like under reasonably favorable historical conditions that are possible given the laws and tendencies of society?" (Rawls 2002, p. 11).

For Rawls, nonideal theory measures the gap between ideal (strict) and actual (partial) compliance, and proposes public policy solutions intended to close that gap. Strict compliance, then, is one of the features that not only distinguishes ideal from nonideal societies, but also differentiates the tasks of ideal and nonideal theory. Ideal theory sets for itself the goal of describing a utopia in which strict compliance is realistically probable. The task for nonideal theory is to suggest ways to achieve that utopia. Rawls acknowledges the importance of nonideal theory in addressing those problems that, "are the pressing and urgent matters…that we are faced with in everyday life" (Rawls 1999, 8). However, he argues that we must begin with ideal theory because, "it provides…the only basis for the systematic grasp of these more pressing problems" (Rawls 1999, 8). Thus, Rawls would identify nonideal conditions as any that fail to measure up to the ideal, placing ideal theorizing in a temporally prior position. Nonideal theory, then, represents a response to injustices identified by the principles of justice that are the outcome of ideal theorizing.

Rawls argues that the principles of justice arrived at in the original position must be further measured against our considered moral judgments, those judgments we hold about justice and morality that we have critically and rationally scrutinized and deem reliable. Reflective equilibrium, his method for doing so, is a checkpoint to assess whether the principles chosen in the original position fit our considered judgments, or reveal flaws in the form of doubts and hesitation, about those judgments (Rawls 1999, p. 17). Considered moral judgments that once seemed plausible may, when examined in the light of the principles chosen in the original position, seem less than plausible. Rawls uses slavery and religious intolerance as examples. While it may be possible for a slave owner to rationalize the value of slavery, behind a veil

of ignorance in which such an individual is unaware of his or her own social and economic position, the slave owner will likely prefer the impermissibility of slavery to ensure that he or she will never suffer such an injustice. In this instance, when the veil is lifted and reflective equilibrium is engaged in, former convictions about the permissibility of slavery would require adjustment. Rawls writes,

> By going back and forth, sometimes altering the conditions of the contractual circumstances, at others withdrawing our judgments and conforming them to principle, I assume that eventually we shall find a description of the initial situation that both expresses reasonable conditions and yields principles which match our considered judgments duly pruned and adjusted. This state of affairs I refer to as reflective equilibrium. It is an equilibrium because at last our principles and judgments coincide; and it is reflective since we know to what principles our judgments conform and the premises of their derivation (Rawls 1999, p. 18).

Thus, reflective equilibrium proposes that we scrutinize the principles of justice to see if they (a) fit what we conceive to be just, or (b) point to a need for modification, leaving open for further reflection whether such modification ought to be applied to the principles of justice or to our considered judgments. When no further modifications are deemed required we have achieved an acceptable, albeit narrow, balance or equilibrium between our principles and our considered judgments.

The back and forth process of reflective equilibrium, Rawls predicts, will result in theoretical principles of justice that conform to ideals of fairness and equality and provide insight into our basic beliefs and assumptions about justice. Once reflective equilibrium is achieved, it will either reveal that our principles confirm our considered moral judgments or it will help us balance intuitions whose weights are unclear and, "…offer a resolution which we can affirm on reflection" (Rawls 1999, p. 17).

Rawls' theory of justice as fairness provides some of the

groundwork for further investigation into differentiating the functions of ideal and nonideal theory in identifying and analyzing injustices as they occur in society, which I will discuss in the next section.

Section 2

Before summarizing the respective views of Simmons and Mills, it is important to distinguish two senses of priority. Temporal priority identifies which type of theorizing (ideal or nonideal) must logically be done first. Rawls and Simmons claim that we must begin with ideal theory. Efficacious priority identifies which type of theorizing is more effective in meeting specific challenges. Mills specifically challenges the ability of ideal theory to contribute substantially to identifying principles of justice representative of an overlapping consensus. The issues of temporal and effective priority, thus, define the context within which ideal and nonideal theorists joust for position in the arena of political philosophy. The objective of this paper is to propose an alternative to prioritizing the roles of ideal and nonideal theory.

John Simmons defends Rawls against the common critical complaint, "that our principal focus in theorizing about justice should be on the part of a theory of justice where, as it were, the rubber meets the road, namely, on nonideal theory" (Simmons 2010, p. 30). He argues that Rawls correctly identifies the priority of ideal theory, and supports Rawls' response to similar criticisms that, "until the ideal is identified…nonideal theory lacks an objective, an aim, by reference to which its queries can be answered" (Rawls 2002, p. 90). Characterizing the priority objection in temporal terms, Simmons argues, like Rawls, that ideal theory lays the groundwork that identifies the task and content of nonideal theory, and therefore is foundational. Alternatively, the most extreme argument for the priority of nonideal theory suggests, "that we can do without the ideal theory of justice altogether. The assumptions…are so far removed from the actual circumstances in which we find ourselves that they are more or less useless in

generating guides for practical reasoning" (Simmons 2010, p. 31).

To the criticism that the assumptions of ideal theory, such as strict compliance and ideal historical conditions, are counterfactual, Simmons responds, "It is precisely the job of nonideal theory to measure that distance [between ideal and current conditions] and to prescribe courses of action that are sensitive to real (i.e., nonideal) world facts…" (Simmons 2010, p. 32). He suggests that the real objection is not that ideal theorizing is useless, but that Rawls' ideal theory ignores common social injustices. He argues that changing the content of ideal theory is not fatal to Rawls' ideal-nonideal distinction, but defends Rawls' insensitivity to injustice based on the fact that, "the correct conception of justice must be a 'political' conception—one that is 'stable for the right reason,' the possible object of an 'overlapping consensus' of competing comprehensive conceptions—implies that there can be no true political community in a modern, pluralistic society without tyranny and injustice (*PL*, pp. 37, 146; *JAF*, pp. 3, 21)" (Simmons 2010, p. 33).[1] Simmons identifies Rawls' response to the temporal priority issue as "importantly correct" (Simmons 2010, p. 34). Without ideal theory, nonideal theory "lacks an objective" (Rawls 2002, p. 90). Thus, for Simmons, specifying the ideal is prior to theorizing the nonideal and how to achieve the ideal. Importantly for my later argument, he claims that it is unnecessary for nonideal theorists to wait for ideal theory to be completed, or even further refined, in order to begin their own work. "To insist only that ideal theory have this kind of priority and this kind of role in informing our political activism in a nonideal world is surely not enough to undermine the cause of justice, but only to insist that we be careful with our political sledgehammers and seek justice thoughtfully" (Simmons 2010, pp. 35–36).

Alternatively, Charles W. Mills claims that nonideal theory is more effective in identifying injustice,[2] arguing that ideal theory, "can be thought of as in part *ideological*, in the pejorative sense of a set of group ideas that reflect, and contribute to perpetuating, illicit group privilege" (Mills 2005, p. 166).

Mills argues that a non-idealizing approach to ethical theory

has the potential to address injustices overlooked by Rawls' principles of justice.[3] He distinguishes ideal theory from nonideal theory not by strict compliance, as do Rawls and Simmons, but by its "reliance on idealization to the exclusion, or at least marginalization, of the actual" (Mills 2005, p. 168). Mills' criticism of ideal theory is that it "either tacitly represents the actual as a simple deviation from the ideal…or claims [as do Rawls and Simmons] that starting from the ideal is at least the best way of realizing it" (Mills 2005, p. 168). Joining other critics, Mills identifies oppression as one of the concepts that ideal theory overlooks, and claims that Rawls' strict compliance assumption reflects "deeper structural biases in the profession" (Mills 2005, p. 169). Mills claims that the failure to explore the significance of the deep difference between idealization and non-idealization guarantees that the idealized will never be achieved, i.e., that Rawls' utopia is not realistic and that ignoring the interests of subjugated groups cannot possibly serve their interests. He concludes that "ideal theory can only serve the interests of the privileged, who, in addition—precisely because of that privilege (as bourgeois white males[4])—have an experience that comes closest to that ideal, and so experience the least cognitive dissonance between it and reality…" (Mills 2005, pp. 170, 172).

Section 3

Simmons' explication of Rawls' distinction between ideal and nonideal theory offers a richly detailed account of what he takes to be Rawls' position "on the relationship between and the content of ideal and nonideal theory" (Simmons 2010, p. 6). Some sense of an ideal must precede any identification of circumstances that fail its requirements, and Rawls and Simmons correctly identify ideal theory as a fundamental starting place in political philosophy. However, this is true only if ideal theory is successful in identifying injustices whose elimination are the objective (as articulated by ideal theorists) of nonideal theorists, and it is not. Importantly, as I will discuss later, Simmons further allows for the work of

nonideal theory to proceed before ideal theory is settled (Simmons 2010, p. 36).

What is lacking in the above articulation of the division of labor between ideal and nonideal theorists is an understanding of how ideal and nonideal theory inform one another. A more thorough analysis of the ideal-nonideal distinction requires, in particular, an account of the way in which nonideal theory informs ideal theory. Simmons' claim is that nonideal theory can get to work on closing the gap between the settled portion of ideal theory and current conditions before it has access to the full ideal. This narrows the portion of ideal theory with which nonideal theory interacts to that which is already settled. That some conception of the ideal is always at the forefront of nonideal speculation has been well articulated by both Simmons and Rawls. However, Simmons' assessment fails to articulate the way that the structure and not just the content of ideal theory is importantly informed by nonideal theorizing. To wit: the inclusion of the insights of nonideal theorists in identifying and analyzing injustices that ideal theory overlooks. Two important oversights have been identified, as well, by Schwartzman and Okin. In her discussion of rights, Schwartzman discusses the claim quoted earlier, "that although rights can be used in arguments for women's equality, they can also function to uphold the power of privileged groups" (Schwartzman 2007, p.15). Like Mills, Schwartzman recognizes that there are ways in which the articulated ideals of equality can favor some groups while obstructing the efforts of others to obtain a similar equality within a liberal democracy whose foundational ideology touts equality for all. This deficiency is especially egregious where principles of justice fail to identify injustices that might otherwise be rectified through the efforts of nonideal theorists. While Schwartzman and Mills focus on ideal theory's failure to attempt to identify the nonideal (injustice) within the context of ideal theorizing, Susan Okin adds that Rawls' separation of public and private spheres woefully neglects the specific subject of gender oppression, relegating women to a sphere of society that is under-theorized by ideal theorists. I am grateful to

Okin for highlighting what I would deem a significant deficiency in any re-characterization of the original position that ignores the fact that "the potential significance of feminist discoveries and conclusions about gender for issues of social justice…undermine centuries of argument that started with the notion that not only the distinct differentiation of women and men but the domination of women by men, being natural, was therefore inevitable and *not even to be considered in discussions of justice* [emphasis added]" (Okin, 1989, p. 7).

The traditional goal of ideal theory is to identify ideal principles of justice. Conditions that fall short of such principles are, thus, identified as unjust and placed within the context of an objective for nonideal theorizing. However, since nonideal theory capably identifies injustices, particularly those that are not identified by even 'settled' ideal theory, the scope of its task seems too narrow where it is limited to simply closing the gap between the ideal and the actual. It is especially important that ideal theorists avail themselves of the unique perspective and ability of nonideal theorizing to analyze the mechanisms of oppression. In this way, ideal theorists may address improvements in the principles of justice through a revision process already provided for by Rawls' reflective equilibrium. It is my contention that nonideal theory identifies problems that importantly improve ideal theorizing about principles of justice, especially the mechanisms by which strict compliance allows injustices to slip through the cracks. So, while ideal theory sets the objective to which nonideal theory must respond, nonideal theory reveals defects in the objective itself to which ideal theory must respond. Thus, a highly productive symbiotic relationship between the two is established in which ideal and nonideal theory work together and in which neither can claim to be more fundamental than the other; that relationship is described by Rawls' method of reflective equilibrium.

The identification of any circumstance deemed, intuitively, to be less than ideal entails a vision of some ideal. The fact that nonideal theorists are able to identify injustices not identified by ideal theory proper indicates not only that ideal theory is not yet

fully informed with respect to an accurate and detailed account of a realistically ideal utopia, but also that nonideal theory can provide some of the answers. Thus, input from nonideal theory is critical to ideal theory if it identifies portions of ideal theory that have failed to recognize a realizable ideal of justice. Certainly injustice has no place in the vision of an ideal society. A society in which citizens and institutions strictly comply with principles of justice and are still oppressed and unjustly dominated is not an ideal society at all, and a vision of a society in which all injustices are at least recognized in order to be addressed, even if unsuccessfully, is required of any theory claiming to be ideal. From this perspective, ideal theory may never be completely settled. As it proceeds towards a fully informed articulation of principles of justice, I believe ideal theory must be flexible enough to incorporate valuable, new insights from nonideal theorists. If ideal theory is incapable of any effective response to improvements in the ability of nonideal theory to pick out injustice, then its attempts to frame a *realistic* utopia seem doomed to failure.

In the opening paragraph of this paper, I argued that trust withers where injustice thrives, weakening the ties that hold a society together. I am now ready to claim that the strength of this social fabric, the purpose for which principles of justice are derived, can be increased by insights made available by nonideal theorists for improving the outcome of Rawls' original position through reflective equilibrium.

Part of the elegance of Rawls' theory of justice is captured by the insight of reflective equilibrium. Far from claiming that the principles of justice are settled, Rawls proposes that they must match our considered judgments. While a discussion of the wisdom of using considered judgments in reflective equilibrium is beyond the scope of this paper, I am convinced of the merit of Rawls' proposal, and will proceed on the assumption that it is worthwhile.

Rawls fortifies his theory of justice with a method he calls "reflective equilibrium" for assessing principles of justice. This makes room for modification of the characterization of the orig-

inal position should the principles of justice which are its outcome require adjustment in order to match our considered judgments. As I have been arguing, the considered judgments of nonideal theorists indicate that some adjustment in the principles of justice may be required in order to describe a utopia that is realistically just.

The concept of a utopia entails a desire to improve current conditions. Yet, the desire to improve, as Simmons has pointed out, is generated by some sense of an ideal, the basic intuitions we all have about justice. Because these same types of intuitions are also available to nonideal theorists, a back and forth working of ideal and nonideal together towards a more realistic utopia appears to be a more inclusive approach to theorizing about justice. While basic, common sense intuitions about what is just drive comparisons between ideal theory and current conditions that identify some injustices, the fact that such intuitions are available to both ideal and nonideal theorists enables Mills to claim that nonideal theory offers insights which ought to be utilized to inform a more realistic ideal. An individual's concept of justice is, he argues, biased by experience. In the case of modern political philosophy, the experience most often applied to justice is that of, "middle-to-upper-class white males—who are hugely *over-represented* in the professional philosophical population" (Mills 2005, p. 172). The fact that this bias has been identified by nonideal theorists suggests that structural changes for which Rawls has made provision in reflective equilibrium are called for in order for ideal theory to be representative of an overlapping consensus. Moreover, the identification of bias in ideal theory by nonideal theorists implies the need for a closer relationship between ideal and nonideal theorizing than what Rawls and Simmons seem to advocate. Any attempt to envision an ideal society without a comprehensive understanding of the mechanisms of nonideal realities, like relations of domination, cripples the ability of ideal theory to guide nonideal theorists towards a resolution, which, according to Rawls, is the task of ideal theory (Rawls 1999, p. 8).

While Simmons concedes that the content of ideal theory

can change, neither he nor Rawls has articulated a willingness to allow for the possibility that input from both ideal and nonideal theory is critical to an articulation of a realistic and fully informed ideal that is resistant to compatibility with injustice.

SECTION 4

Lisa Schwartzman, among others, has identified something upon which I will elaborate upon and call the strict compliance problem.[5] A quote from Rawls will help:

> I want to examine the principles of justice that would regulate a well-ordered society…thus I consider primarily what I will call strict compliance as opposed to partial compliance theory. The latter studies the principles that govern how we are to deal with injustice… It comprises such topics as the theory of punishment, the doctrine of just war [etc.]…Obviously the problems of partial compliance theory are the pressing and urgent matters. These are the things that we are faced with in everyday life. The reason for beginning with ideal theory is that it provides, I believe, the only basis for the systematic grasp of these more pressing problems" (Rawls 1999, p. 8).

In the above passage, Rawls uses "strict compliance theory" and "ideal theory" interchangeably. If strict compliance denotes the ideal, and partial compliance denotes the nonideal, a logical move is to use strict compliance as a measure of the gap between the ideal and the actual that nonideal theorists are charged with closing. However, there are two problematic outcomes if strict compliance to principles of justice is compatible with injustice, as Schwartzman claims.[6] Either (a) the ability of principles of justice to correctly distinguish between justice and injustice fails, resulting in the inability of ideal theory to set a proper objective for nonideal theory, or (b) the pragmatism of strict compliance as a measurement is called into question. Although a third possibility does exist, that one must concede, as does Rawls, that some

injustice is inevitable, I want to claim that such injustices must at least be recognized and posed as an objective before nonideal theorists, and ideal theory fails to articulate any responsibility to do so. While Rawls' position is incompatible with the very idea of a utopia, it would nevertheless be consistent with a *realistic* utopia, which is the context of this paper. However, a concession that some injustice is inevitable does not relieve the pressure to isolate instances of injustice and theorize their sources and methodologies. This pressure arises from the aspiration to eradicate as much injustice as possible in order to achieve the realistic utopia Rawls later identifies in *The Law of Peoples*.

Both Rawls and Simmons identify strict compliance to principles of justice as the measure of a well-ordered society. Accordingly, the implication is that the measure of strict compliance to such principles identifies and distinguishes just from unjust institutions, measuring whether those institutions meet the requirements of justice as outlined in the principles of justice. Thus, an institution that is judged to be in strict compliance with principles of justice is deemed a just institution. An institution judged not to be in strict compliance, or one that is in partial or even noncompliance, is judged to be unjust or simply not as just as strict compliance requires. Finally, the gap between strict and partial compliance (or noncompliance) is the measure by which nonideal theory determines appropriate policy; that is, policy intended to close the gap. Any compatibility between strict compliance and injustice (that is, any institution that can pass the strict compatibility requirement while, at the same time, allowing for injustices to exist within its own sphere of influence) necessarily signals a problem with the ability of strict compliance to serve either as a measure of ideality or as a guide for nonideal theory. Thus, if strict compliance is compatible with oppression and "illicit group privilege," as Schwartzman and Mills claim, then not only has ideal theory failed to identify a realistically just society, but it has failed to provide nonideal theory with an adequate objective that Rawls and Simmons argue is its purpose (Schwartzman 2007, p. 15; Mills 2005, p. 166). The notion that institutions are just as

long as they are in strict compliance with principles of justice is a meaningless measurement of justice if injustices are allowed to prevail within the parameters of that determination.

The strict compliance problem is generated either by principles of justice that fail to meet the basic requirements of justice to begin with or by the ability of strict compliance to pick out injustice. It seems reasonable to assume that nonideal theorists have access to the same intuitions and considered judgments about justice as do ideal theorists when characterizing the original position, yet nonideal theorists have demonstrated that they are able to identify injustices that strict compliance to principles of justice allows. This suggests that there is a problematic disconnect between the intuitions and considered judgments that ideal theorists use to characterize the original position, and the intuitions and considered judgments that nonideal theorists use in theorizing oppression and group privilege. While it is unclear what, exactly, comprises this disconnect, it is clear that nonideal theory has something valuable to contribute to ideal theorizing. Nonideal theory, then, can make a substantial contribution to the process of determining whether the principles of justice for which ideal theory assumes strict compliance in a perfectly just society are complete or incomplete in a way that ideal theorists have, so far, failed to do.

There are two broad sets of possible objections to the original position argument that threaten the usefulness of the principles of justice in accurately assessing the justness of social institutions. The first objection is that Rawls' characterization of the original position allows too much, thwarting the intent to filter out bias, or information that would allow the parties to tailor principles to favor their own situation. This is the broad domain of Mills' objections. The second objection is that it doesn't allow enough. The exclusion of certain historical data results in theorizing that is too far removed (a) from realistic human inclinations and abilities, and (b) from factual historical circumstances that are, unlike Rawls' reasonably favorable historical circumstances, not reasonable at all. The second (exclusion) objection results in the claim

that ideal theory is unable to generate principles of justice that distinguish the just institutions that are characteristic of Rawls' realistic utopia.

In the original position, representative persons contract for principles of justice that will eventually guide the ensuing social structure and its institutions. They do so under a veil of ignorance that is intended to exclude bias and favor fairness. Arguing "… from widely accepted but weak premises…," Rawls' restrictions are intended to enable the parties to reflect upon those hypothetical choices they might make for principles of justice under less individually biased conditions (Rawls 1999, p. 16). His argument for the veil of ignorance, then, is that it ensures a higher degree of impartiality in how the principles are chosen than would otherwise be achievable due to biases of human nature that encourage each man to place his own interests above those of others. However, not all bias has been excluded. Importantly, as evidenced by the arguments of Mills, Schwartzman and Okin, relations of domination continue to exacerbate the problems Rawls would like to eliminate[7] even when institutions qualify as just under an assumption of strict compliance. This problem can be at least partly resolved by including factual historical data, data that Rawls has inadvertently mischaracterized as favorable, in a re-characterization of the original position in reflective equilibrium. The inclusion of factual historical data is necessary to a re-characterization of the original position that anticipates principles of justice that speak more fully to, and are able to isolate, oppression and relations of domination. Strict compliance to Rawls' principles cannot filter out such relations, and ideal theory has thus far been unable to identify such relations in order to motivate such a re-characterization. It is nonideal theory that has been able to isolate the mechanisms of relations of domination and, thus, nonideal theory that informs ideal theory at least as much as ideal theory informs the nonideal.

My claim is that the utopia described by the principles of justice for which Rawls argues is not yet realistic because it allows for injustices whose eradication may be out of reach simply because they have not yet been recognized. I have argued that institutions

complying with Rawls' principles of justice allow for injustice to prevail where it should not, and that his principles do not provide for an adequate identification of injustices that is required in order to theorize problems arising out of partial or noncompliance. I have offered a solution by analyzing the intersection of ideal and nonideal theory, comprised of an overlapping focus on the injustice of oppression and domination, which I believe will promote a productive exchange. Finally, marking out a space where theorists engaged in both ideal and nonideal political philosophy might work together, reflective equilibrium provides within that space the possibility of an articulation of principles of justice more complete than either has thus far achieved alone.

Notes

1, Simmons' references are to Rawls' Justice as Fairness (*JAF*), and Political Liberalism (*PL*).

2, As mentioned in the introduction, Mills argues that ideal theory, "can be thought of as in part *ideological*, in the pejorative sense of a set of group ideas that reflect, and contribute to perpetuating, illicit group privilege."

3, Specifically, ideal theory's "perpetuation of illicit group privilege."

4, Mills is addressing the problem of overrepresentation of privileged white bourgeois males in the field of philosophy.

5, Schwartzman cites Catharine MacKinnon and Andrea Dworkin in *Pornography and Civil Rights* "rather than winning equality for oppressed minorities, rights often function in ways that uphold current power structures."

6, "although rights can be used in arguments for women's equality, they can also function to uphold the power of privileged groups" (Schwartzman 2007, p.15).

7, "unjust war and oppression, religious persecution and the denial of liberty of conscience, starvation and poverty,…genocide and mass murder" (Rawls 2002, p. 7).

Bibliography

Korsgaard, Christine M. (1986) "The Right to Lie: Kant on Dealing with Evil," *Philosophy and Public Affairs* Vol. 15, No. 4, pp. 325–349.

Mills, Charles W. (2005) “Ideal theory as Ideology,” *Hypatia* vol. 20, no. 3, pp. 165–184.

Okin, Susan Moller. (1989) *Justice, Gender, and the Family* (Basic Books: United States of America)

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

_____. (2002) *The Law of Peoples* (Cambridge: Harvard University Press)

Robeyns, Ingrid (2008) “Ideal Theory in Theory and Practice,” *Social Theory and Practice* 34, pp. 341-362.

Schwartzman, Lisa H. (2007) *Challenging Liberalism: Feminism as Political Critique* (University Park: Pennsylvania State University Press)

Simmons, A. John (2010) “Ideal and Nonideal Theory,” *Philosophy & Public Affairs* 38, No. 1, pp. 5–36).

"Outing" Power: Exposing evasion and concealment of power through "sexuality" in Foucault's *History of Sexuality Vol. 1*

Anthony Ristow

"Finally, the notion of sex brought about a fundamental reversal; it made it possible to invert the representation of the relationships of power to sexuality, causing the latter to appear, not in its essential and positive relation to power, but as being rooted in a specific and irreducible urgency which power tries as best it can to dominate; thus the idea of 'sex' makes it possible to evade what gives 'power' its power…" (Foucault 1978, *HS*, p. 155).

Michel Foucault's texts are no doubt challenging, both in complexity and social implication. They can also be perplexing. Yet one finds oneself, as Ladelle McWhorter did, "wickedly delighted by his disruption of established order" (McWhorter 1999, *B&P*, p. 62). His work deconstructs and problematizes accepted knowledge that is comfortably viewed and given as natural—like the concept of "sexuality"—and exposes it as a product of strategies, thoroughly and inextricably enmeshed in its relationship to unstable power. Something as familiar as our sexuality is undergoing a disconcerting reversal, but the possibilities Foucault's thought opens up can be profoundly liberating.

In *The History of Sexuality Vol. 1*, Foucault does a genealogical critique of the emergence and development of "sexuality" as a concept over the last three centuries in the West and the subsequent "repressive hypothesis" which still conditions how we understand and behave towards sex today. Sex, for Foucault, is not an irreducible, universal given. We think we all possess

this fundamental part of ourselves called sex but it is "precisely this idea of sex *in itself* that we cannot accept without examination" (Foucault 1978, *HS*, p. 151). In his examination, Foucault sees in us "the elaboration of this idea that there exists something other than bodies, organs, somatic locations, functions, anatomo-physiological systems, sensations, and pleasures; something else and something more, with intrinsic properties and laws of its own: 'sex'" (Foucault 1978, *HS*, p. 151). That is not to say that Foucault denies the existence of sex organs or sex drives before 1800 but rather that they were organized and construed in a new and particular way since the late nineteenth century we call sexuality. Sexuality, Foucault argues, is not the expression of sex but as McWhorter interprets, "[s]exuality is a regime of power and knowledge that develops within institutions and practices that aim to harness the strength and developmental potential of human bodies and put them to use…" (McWhorter 2004, p. 40). That is to say sex is not an essential part of us seized upon and dominated by power as we have come to think of it. Rather, sex is an idea that was actually "formed inside the deployment of sexuality," a product of this power-knowledge and "historically subordinate" to it (Foucault 1978, *HS*, p. 152). In other words, certain organs, functions, and pleasures have been grouped together into an "artificial unity" called sex that serves as an imaginary anchor point and target for the regimes of power Foucault dubs the *dispositif de sexualite*, or deployment of sexuality.

According to Foucault, this deployment implanted sex with a secret—a truth revealing function—tied to the resounding belief that we are sexually repressed, and to such an extent "that we became dedicated to the endless task of forcing its secret, of exacting the truest of confessions from a shadow" (Foucault 1978, *HS*, p. 159). The irony of this great deployment of sexuality that occurred in Western society, particularly from the eighteenth century onward, is that it would lead us to believe our "liberation" of all things was at stake in the imperative to make sex speak.

My contention is that articulating what is incendiary about this text hinges on Foucault's concept of power as producing sex.

Specifically, I am interested in how power manifests and proliferates in the bourgeois theory of repression, and in that theory's relationship to the idea that "power is tolerable only on condition that it mask a substantial part of itself" (Foucault 1978, *HS*, p. 86). In this essay, I pose two questions: What does Foucault mean when he argues that power is successful to the extent that it is able to conceal its operations? And if power operates by masking itself, what kind of resistance would be effective? I will argue that a purely negative and prohibitive account of power in terms of a natural, irreducible sex is not only a misapprehension; it also enables power to remain concealed. Otherwise, why would we consent to it, let alone search within its grips for a liberation that not only does not exist, but also actually helps to proliferate our own limited view of sex in the process?

This essay is broken up into three parts. I begin with a preliminary look at Foucault's conception of power followed in part two with a discussion of the "repressive hypothesis" in four sections. The first two sections will explicate what exactly Foucault means by the "repressive hypothesis" and how repression became the principal way we have come to misunderstand power and sexuality. I will follow this in section three by explaining how such a misunderstanding has implanted sexuality with a liberation imperative that not only does not liberate us as sexual subjects but reinforces established centers of power and further enables our sexual limitation. I argue in the final section of part two that this liberation project is still rampant and pervasive in contemporary sexuality today. I will attempt to illustrate the "repressive hypothesis" at work in contemporary American sexuality with concrete and local examples from Ariel Levy's recent pop culture study, *Female Chauvinist Pigs*. My hope in this fourth section is not only to make a Foucauldian critique of sexuality more accessible to the uninitiated but to also show why it is extremely relevant today. Foucault's unique analysis of power and critical account of our sexuality is not limited to the generations following the Victorian era but is perhaps even more crucial for the descendents of sexual "revolution."

With part three, I make an interpretive move connecting the concealment of power's devious and productive mechanisms—and the integral role disguise plays—to the success of the "repressive hypothesis." And in conclusion, I will argue that resistance lies in "outing" power. The operations of power permeating sexuality can be revealed in two ways: first, by diagnosing its hidden mechanisms and second, by exposing the phony notion of "sexual liberation." In order to perform an analysis of how power operates through sexuality, however, I must first give a general overview of how Foucault conceives of power.

PART ONE:
METHOD OF *POWER*

"One needs to be nominalistic, no doubt: power is not an institution, and not a structure; neither is it a certain strength we are endowed with; it is the name that one attributes to a complex strategical situation in a particular society" (Foucault 1978, *HS*, p. 93).

One finds Foucault's most direct and extensive discussion of power in the "Method" chapter of *The History of Sexuality*. Foucault does an analysis of knowledge regarding sex not in the juridical sense, as mandated by law, nor as something that is repressed, but as a product of *power*. For Foucault, power is an exercise—"not something that is acquired, seized, or shared," not something one gains or loses (Foucault 1978, *HS*, p. 94). In other words, power is not a tangible possession; it is a process of relational forces. However, power relations cannot be thought of simply in the traditional terms of master/servant or dominant/dominated. Power must be understood as a "multiplicity of force relations" which emerge "through ceaseless struggles and confrontations" that are "immanent in the sphere in which they operate." Most importantly, power "is the moving substrate of force relations which, by virtue of their inequality, constantly engender states of power, but the latter are always local and

unstable" (Foucault 1978, *HS*, p. 92–93). Like a volatile language, these force relations that constitute Foucault's idea of power are not static and can organize themselves in such a way as to form constantly changing systems of meanings or values. Their "grid of intelligibility" is determined by contingent and arbitrary differences and cleavages rather than anything fundamental.

"Power" is the name Foucault gives strategies and mechanisms in a particular society; but there are no definitive authors behind the movements and permutations of society—this is what Dreyfus and Rabinow aptly call "strategies without strategists" (Dreyfus & Rabinow 1983, *MF*, p. 109). This conception of power has an impersonal nature; it only emerges through tactics, practices, and "meticulous rituals" rather than individual agents harnessing the separate entity that is power and exercising it towards their means (Dreyfus & Rabinow 1983, *MF*, p. 109). Whereas conventional thought might conceive of pre-existing "substantial entities," e.g. subjects, as the condition for a relationship of struggle from which power is produced, Foucault sees these entities as created by the very emergence of that struggle of forces. They are bound together as a creation of performance or exercise. Put another way, "The world is not a play which simply masks a truer reality that exists behind the scenes. It is as it appears. This is the profundity of the genealogist's insight" (Dreyfus & Rabinow 1983, *MF*, p. 109). It is as it appears, apparent and immanent in our practices, yet we mask that superficial reality by telling ourselves that the most important, the most fundamental aspects are murky secrets and hidden meanings that must be sought out and interpreted. Through a peculiar will to knowledge and perhaps the hubris of self importance and affirmation, we make plain truth about power unintelligible in favor of something akin to buried treasure.

To further clarify, Foucault lays out four rules to follow regarding power. These are not "methodological imperatives" but can serve as "cautionary prescriptions." First, the *Rule of immanence*: "Between techniques of knowledge and strategies of power, there is no exteriority…" In other words, they are inex-

tricably bound together (Foucault 1978, *HS*, p. 98). Second, the *Rules of continual variations*: We must not look in terms of who has the power and who is deprived because power relations and their distribution are variant, undergoing continual shifts (Foucault 1978, *HS*, p. 99). Third, the *Rule of double conditioning*: All "local centers" of power are part of an "over-all strategy" and larger strategies rely on "local centers" as an "anchor point;" they play and subsist off each other (Foucault 1978, *HS*, p. 99). And fourth, the *Rule of the tactical polyvalence of discourses*: Discourse "can be both an instrument and effect of power, but also a hindrance… a point of resistance and a starting point for an opposing strategy." That is to say, "there can exist different and even contradictory discourses within the same strategy" (Foucault 1978, *HS*, p. 101–102).

Foucault ends his address on power in *The History of Sexuality* by emphasizing that his is a strategical model rather than one based on law alone; a conception "which replaces the privilege of the law with the viewpoint of the objective, the privilege of prohibition with the viewpoint of tactical efficacy, the privilege of sovereignty with the analysis of a multiple and mobile field of force relations, wherein far-reaching, but never completely stable, effects of domination are produced" (Foucault 1978, *HS*, p. 102). And this point is important because it not only indicates a productive concept of power over a prohibitive one, and how a multitude of localized strategies are a boon and relay for a greater deployment of domination, but also that these forces are mobile and unstable, which allows for the possibility that they can be changed.

PART TWO:

I. The "Repressive Hypothesis"

"The question I would like to pose is not, Why are we repressed? But rather, Why do we say, with so much passion and so much resentment against our most recent past, against our present, and against ourselves, that we are repressed?" (Foucault 1978, *HS*, p. 8)

For Foucault, the well-worn idea that we in Western society have been sexually repressed since the Victorian era—the image of the hypocritical, "imperial prude"—is a dubious hypothesis. Foucault calls this the "repressive hypothesis." On the contrary, he demonstrates that a marked "explosion in discourse" on the matter of sex shows at least that it is, and has been, a prevalent and abundant obsession in Western society and not the taboo subject sentenced to nonexistence and silence the repressive hypothesis would have us believe. The running narrative has been that sex once enjoyed a more open and far brighter day now long gone in the twilight of the "monotonous nights of the Victorian bourgeoisie;" repressive forces have relegated sex to the parents' bedroom, concerned primarily with the "serious function of reproduction" (Foucault 1978, p. 3). Foucault critiques the idea that the locus of state-approved sex was placed on the conjugal family for purposes of maximizing labor power in the burgeoning capitalism of the West while repressing all other forms of sexual relations and energies. "If sex is so rigorously repressed, this is because it is incompatible with a general and intensive work imperative" (Foucault 1978, p. 6). They (the bourgeois storytellers of Western sex) placed the advent of the age of repression in the seventeenth century at a time when labor capacity was at a premium, being systematically exploited, so it couldn't be allowed to waste itself on pleasure. If this were the case one would think the young adult, proletariat male, possessing nothing but his life force, "had to be the primary target of subjugation destined to shift the energy available for useless pleasure toward compulsory labor" (Foucault 1978, p. 120)—that his *bio-power* would be the ideal and first to be harnessed by the capitalist powers that be. However, Foucault finds that, "on the contrary, the most rigorous techniques were formed and applied first, with the greatest intensity, in the economically privileged and politically dominant classes first." One can assume only the bourgeoisie could afford the leisure time and money to devote to psychiatry and the original problematization of sexual development. For example, some of the first subjects to

be "sexualized" were the "idle" woman and the onanistic child—not a child of the people, a future worker, but a schoolboy with servants and tutors (Foucault 1978, p. 121).

Foucault argues that the preceding explanation for how and why our sex supposedly got so repressed was *adjusted* to coincide with the development of capitalism, thoroughly integrating it into the "bourgeois order" (Foucault 1978, p. 5). This account serves as a post facto explanation that oversimplifies the more challenging and accurate analysis of what has happened to sex in the Victorian era. Foucault believes the techniques of power have not restricted sex in favor of labor capacity, but produced a proliferation, incitement, and implantation of a multitude of sexualities. This proliferation first occurred in the bourgeois class and eventually trickled down to the working class, not in accordance with some "Marxist" account of repression, but due to a class-motivated production as we shall see in greater detail in the following section. Once the concepts of a body and of a sex were "finally conceded to them," the working class was subsequently placed under surveillance through various techniques (whether in schooling, housing politics, public hygiene, the general medicalization of the population, etc.) (Foucault 1978, p. 126). But these techniques first developed along with an exigent "pathology of sex" and its techniques of maintenance and correction in the aristocratic family: "The bourgeoisie began considering that its own sex was something important, a fragile treasure, a secret that had to be discovered at all costs" (Foucault 1978, p. 120–121).

II. The Discreet Charm of Bourgeois Sexuality

"It provided itself with a body to be cared for, protected, cultivated, and preserved from the many dangers and contacts, to be isolated from others so that it would retain its differential value…" (Foucault, *HS*, p. 123).

According to Foucault, the ruling classes first tried out the

great sexual deployment on themselves. Contra the repressive hypothesis, sexuality was not deployed by repressive forces but through productive, developmental, and life-affirming incitements and concerns that had exploitative consequences. The Victorians were not concerned with sex from a religious, flesh-denying stance, nor were they concerned with subverting the reproductive forces of populations into a capitalist work ethic. Rather, these techniques intensified the body, problematizing health, and making sexuality a "question of techniques for maximizing life"—in terms of "the body, vigor, longevity, progeniture, and descent of the classes that 'ruled'" (Foucault 1978, p. 122–123). Consequently, the emergence of a biopolitics of sexuality "has to be seen as the self-affirmation of one class rather than the enslavement of another" (Foucault 1978, p. 123).

The bourgeoisie differentiated itself from the lower classes not by dominating through rule of law and saying No to sexual drives, but by affirming and worrying over life; they cultivated a special, unique body and sex, imbuing it with a "mysterious and undefined power" as they placed the onus on proper, normal development and "future welfare" of the body. Foucault calls this "the autosexualization of the body" (Foucault 1978, p. 124). Furthermore, according to Foucault, this sexualization, "starting from a hegemonic center," was eventually granted to the whole social body (Foucault 1978, p. 127). It was at *this* point that the theory of repression emerges, in the nineteenth century, when a new differentiating element was introduced for the "precious sexuality" of the bourgeoisie. A discursive transition took place at the end of the eighteenth century where the discourse that constructed sex as "a valuable element within us that must be feared and treated with respect… was replaced by a discourse which said: 'Our sexuality, unlike that of others, is subjected to a regime of repression so intense as to present a constant danger; not only is sex a formidable secret… but if it carries with it so many dangers, this is because… we have too long reduced it to silence'" (Foucault 1978, p. 128-129). By this account, sex in general harbors fundamental truth and must be monitored, but bourgeois sexuality is

especially unique due to its extreme repression. For Foucault, it is at this juncture in our history, as a result of social distantiality, that psychoanalysis emerges as a phenomenon and the repressive hypothesis begins to condition our sexuality as a discursive liberation project. "Henceforth social differentiation would be affirmed, not by the 'sexual' quality of the body but by the intensity of its repression" (Foucault 1978, p. 129).

III. Exploiting Sex as *The* Secret

"It is pointless to ask: Why then is sex so secret? …In reality, this question, so often repeated nowadays, is but a recent form of a considerable affirmation and secular prescription: there is where the truth is; go see if you can uncover it" (Foucault 1978, p. 79).

In the process of telling ourselves that we are repressed—that power is primarily prohibitive, rather than productive—this repressive script mines a progressive tale underneath what is already there in our sexuality, apparent in our material discourses and practices. Yet there is a will and incitement to knowledge that compels us to look deeper into what already resides out in the open: "people will be surprised at the eagerness with which we went about pretending to rouse from its slumber a sexuality which everything—our discourses, our customs, our institutions, our regulations, our knowledges—was busy producing in the light of day and broadcasting to noisy accompaniment" (Foucault 1978, p. 158).

If we tell ourselves that sex is repressed rather than shining forth in the light of day, "then the mere fact that one is speaking about it has the appearance of a deliberate transgression" (Foucault 1978, p. 6). That is, one believes that they place oneself outside the reach of power: "our tone of voice shows that we know that we are being subversive, and we ardently conjure away the present and appeal to the future, whose day will be hastened by the contribution we believe we are making" (Foucault 1978, p. 6). This "contribution" is just a reinforcement of the status quo since

there is "an institutional incitement to speak about it…" (Foucault 1978, p. 18).

What must be understood, for Foucault, is that when speaking within this narrow framework of the repressive hypothesis, and the deployment of sexuality in general, what one can and will say about sex is a product of, parasitic upon, and exacerbates the channels of power already at work. As long as sex is thought about as solely repressed by power, the will to objective knowledge about sex or even the intention to resist and revolt normative sexual identities in those subjected by the deployment of sexuality is futile. According to Foucault, however, this has not stopped us from trying: "What is peculiar to modern societies, in fact, is not that they consigned sex to a shadow existence, but that they dedicated themselves to speaking of it *ad infinitum*, while exploiting it as *the secret*" (Foucault 1978, p. 35). Foucault's genealogy of sexuality is aimed at destroying the chimera that is breaking the secret of sex and getting underneath to the truth about it. By studying it, inciting some people to speak about it and authorizing others to listen and then interpret with their supposed objective expertise, we are not removing an obstacle but provoking more centers of power to organize around it. The obstacle is a tartuffe, a poser, and a necessary condition of the whole proliferation. A feed-back loop emerges, a tremendous paradox, as sex is portrayed as outside of discourse, and supposedly "abusively reduced to silence," yet there is an incitement to more discourse about it, and "something akin to a secret whose discovery is imperative" through evermore and ever-penetrating discourse (Foucault 1978, p. 35). Foucault calls this "a fable that is indispensable to the endlessly proliferating economy of the discourse on sex" (Foucault 1978, p. 35). What would happen to this proliferating economy and its centers of power if the obstacle were to be removed and this secret was ever to be discovered, or invented…? What would we talk about then?

Foucault is not saying that sexuality has enjoyed (under capitalist and bourgeois Western society) a regime of "unfettered, unchanging liberty;" nor is he saying our societies are more

tolerant than repressive; rather his doubts detect "a more devious and discreet form of power" than a new episode in the lessening of prohibitions can disclose (Foucault 1978, p. 10). It seems what Foucault is ultimately advancing is an analysis of power-knowledge-pleasure (as a technology utilized by agencies of power). He is charting the discursive explosion in sexuality since the late 16th century in Western society, and is diagnosing our failure to understand how networks of power have constructed our perceptions of sexuality. Instead, we have simplistically conceived of sex as an already existing, natural part of ourselves that has merely been seized upon by repressive forces and conservative regimes. Foucault's genealogical work here shows us how these perceptions are not only misapprehended and misunderstood but how they play into the increased monitoring, controlling, managing, and actual production (inception and constitution) of the modern concept, and even a dubious science of, "sexuality," and all that it entails. Our sexualities are not necessarily what we say they are (and Foucault is implying that there are consequences—in the form of normalizing constraints, wasted energy, and bodies laid to waste—for not reckoning with this fact). It certainly follows from this that sex does not harbor a fundamental secret about us. Merely talking about it within the circular plot loop that has constituted the roles it can and will play—while making for peculiar and compelling theatre—does not transgress, subvert, or liberate us from the current power regimes; rather, it perpetuates and proliferates them.

IV. The Irony of this Deployment: Liberation & Empowerment

"We must not think that by saying yes to sex, one says no to power…" (Foucault 1978, *HS*, p. 157).

Through some local and concrete examples of the peculiar sexual drama playing itself out in present-day America, I hope to show how Foucault's critique of our sexual liberation impera-

tive is still valid and relevant today. A brief look at Ariel Levy's journalistic account of some fascinating trends of contemporary sexuality in American popular culture may serve to give a more accessible illustration of Foucault's often elusive and contrarian views regarding our supposed sexual repression.

In *Female Chauvinist Pigs*, Levy does an investigative report on "women and the rise of raunch culture." In Levy's assessment, "[r]aunch culture isn't about opening our minds to the possibilities and mysteries of sexuality. It's about endlessly reiterating one particular—and particularly commercial—shorthand for sexiness" (Levy, *FCP*, p. 30). And this current trend's leading protagonists are what she calls "Female Chauvinist Pigs," or FCPs. "If Male Chauvinist Pigs were men who regarded women as pieces of meat, we would outdo them and be Female Chauvinist Pigs: women who make sex objects of other women and of ourselves" (Levy, *FCP*, p. 5). Levy's main concern is that these FCPs have been branded and marketed under the banner of female empowerment, inciting trends of bawdy sexual expression in the name of rebellion, nonconformity, and liberation when in reality they represent "more a parody of female sexual power than an expression of it" (Levy, *FCP*, p. 98). Levy examines a litany of what she finds to be disturbing new sexual trends in mainstream female sexuality emergent in a drastic cultural shift: "[o]nly thirty years (my lifetime) ago, our mothers were 'burning their bras' and picketing Playboy, and suddenly we were getting implants and wearing the bunny logo as supposed symbols of liberation" (Levy, *FCP*, p. 3). Levy traces this shift back to unresolved conflicts between the women's movement and the sexual revolution. However, my focus in this section is not to do a genealogy of the origins of "raunch culture" but to look at Levy's well documented and concrete cultural examples in order to show an apparent proliferation and intensification of practices that constitute slight variations on the familiar theme of sexual repression and liberation.

Levy's examples range from interviews with college girls who flash their breasts—and often more—in front of the cameras of the hugely successful DVD empire Girls Gone Wild in exchange

for t-shirts and hats; to the emergence of a new porn-chic and a best-selling publishing industry of tell-all memoirs of porn stars. Levy targets the prevalence of breast implants, Brazilian bikini waxes, and other beauty trends—constituting new body norms—as well as thongs marketed to prepubescent girls and the ubiquitous stripper pole and exotic dancers now commonplace if not obligatory in popular music videos. The significance of these examples is that the women and girls Levy calls FCPs do not simply claim to be having fun; nor are they merely clamoring for male attention; they believe they are being brave and funny, battling sexual repression with liberating forms of conspicuous and empowering expression.

While these examples may seem like obvious or easy targets, the fact that they are peddled—and enthusiastically accepted by many—as the solution to some kind of sexual repression (perhaps resulting from the feminist movement or from "political correctness," in Levy's account) makes them a prime example of the insidious mechanisms of power entangled with sexuality that Foucault has in mind. Levy laments that "[t]he proposition that having the most simplistic, plastic stereotypes of female sexuality constantly reiterated throughout our culture somehow proves that we are sexually liberated and personally empowered has been offered to us, and we have accepted it" (Levy, *FCP*, p. 197). She also astutely points out that these forms of sexual expression may actually be empowering for some but within one particular and narrow framework: "sexual power is only one, very specific kind of power. And what's more, looking like a stripper or a Hooters waitress or a Playboy bunny is only one, very specific kind of sexual expression. Is it the one that turns us—*or men*—on the most? We would have to stop endlessly reenacting this one raunchy script in order to find out" (Levy, *FCP*, p. 197-198). She is not going as far as to say there is anything inherently wrong with these expressions in themselves, but to trot them out and flaunt them as legitimate forms of sexual liberation and empowerment—when restricted to and limited within a particular, heteronormative script—is an egregious irony.

My aim is not to pick on the people who buy into this idea of liberation and empowerment. Perhaps something like the new phenomenon of aerobic pole dancing classes may help to positively deconstruct deeply-rooted body-image issues for many women, and in a sense, provide one type of empowerment and speak to a happier existence within the context of this society's normalized sexuality. But in no way does this achieve any new significant sexual freedom. Taking the pole away from the strip club and putting it into all-women spaces, like the hugely popular brand S Factor, or the living room of women's homes (domestic stripper poles are now affordable and widely available) is not a radical act. It does not appropriate or co-opt a source of oppression and redefine it as a point of resistance. If anything, it softens the image and sanitizes the practice for mainstream, normal consumption. If it liberates one from anything, it would be from the constraint of feeling awkward or dirty in enacting one particular, prescribed way of being "sexy," and imparting in the process the capability of performing that one prescription of "sexy" in the safe and acceptable spaces provided. This seems much less like resistance or liberation than assimilation and normalization. Ironically, where the economy of pole dancing once consisted almost exclusively of men paying women to strip in gentlemen's clubs, there now exists a growing subculture of upper-middle class women paying exorbitant amounts of money to learn pole and lap dancing in order to bring these skills home to men—as well as for the purposes of self confidence, empowerment, and fitness. Classes run around $400 per 8 week session in the case of S Factor, roughly $50 per class, for which founder Sheila Kelley promises—in all seriousness—to unleash the "hidden erotic creature," or "center of sexual power and self knowledge" that "exists in every woman" as her "untamed" sexual alter ego—"the opposite of the 'good girl' or 'little lady'" (Kelley, *TSF*, p. 3). Trends of sexual expression such as these, though possibly a good experience for those involved, are obviously thoroughly enmeshed in the power trappings of the great sexual deployment described by Foucault and show that the repressive hypothesis is still rampant today.

I am not interested here in any kind of a traditional, normative moral critique of raunch culture or even in how Levy conceives of true sexual liberation. Rather, what interests me in Levy's investigation is a particular form of popular self-deception based on a gross misunderstanding of power's constitutive effects, parading as a rallying cry against the very limitation its constituents and benefactors claim to resist. In her introduction, echoing Foucault, Levy argues that, in one sense, "[w]hat we once regarded as a *kind* of sexual expression we now view *as* sexuality" (Levy, *FCP*, p. 5). The result seems to be a palatable and recreational version of feminism sanitized of any truly difficult confrontation with deep-rooted societal inequalities. Raunch culture is supposed to be provocative and bawdy fun; yet the subtext and often overriding message of it is not sold as simply fun, but also as a quasi-feminist liberation movement. However, this is a distorted and reassuring fable with consequences that run counter to liberation. Purporting this one trope of sexual expression as the solution to our perceived repression not only implies some straw-man repressive obstacle in our recent cultural past, it also obscures and evades real socio-political problems involving sexuality—i.e., heteronormative sexual oppression, institutionalized patriarchy, etc.—while superficial liberation is just an assimilated variation of what already passes as normal sexual behavior and reinforcement of well established gender roles. As Levy puts it, "[t]he freedom to be sexually provocative or promiscuous is not enough freedom… And we are not free in the sexual arena. We have simply adopted a new norm, a new role to play: lusty, busty exhibitionist" (Levy, *FCP*, p. 200). How can performing what is dictated as sexy by entrenched heterosexist societal norms be transgressive—let alone a feminist solution to the real plight of sexual oppression?

Raunch culture does not say no to (repressive) power by saying yes to a reduced parody of sex. Rather, by "mimicking whatever pop culture holds up to us as sexy" (Levy, *FCP*, p. 200), raunch culture can only impart *feeling* empowered, feeling accepted, feeling more normal within the current context of the power mechanisms of sexuality. In this way, the feeling of freedom

is procured, the progressive illusion of sexual liberation is firmly implanted into the cultural consciousness of repression, and the productive machinations of power are hidden.

PART THREE:
Caché

"…this society that has been more imaginative, probably, than any other in creating devious and supple mechanisms of power, what explains this tendency not to recognize the latter except in the negative and emaciated form of prohibition?" (Foucault 1978, *HS*, p. 68)

Foucault's ultimate concern in *The History of Sexuality Vol. 1* is not to give an exhaustive account of the emergence of Western sexuality; rather, in my reading, his prime aim is to critique a large part of society bent on liberating themselves from sexual repression by means of telling themselves in abundance that they are so sexually repressed, rousing what is in itself its own form of sexuality. As the evidence of this proliferation of sexuality accumulates and convincing indictments mount in the text, the question that starts to arise for the reader, at least for myself, is not so much as why do we tell ourselves with so much passion that we are repressed and how we came to do so, but what, in particular, enables such a grossly misguided understanding of sexuality and power?

By my interpretation, the reason we misunderstand the relationship between power and sexuality can be found in the connection between the repressive hypothesis and what Foucault offers as a "general and tactical reason that seems self evident" but is still profoundly important: "*power is tolerable only on condition that it mask a substantial part of itself*. Its success is proportional to its ability to hide its own mechanisms" (Foucault 1978, *HS*, p. 86). Although not explicitly stated in the text, I argue the most critical application of Foucault's work on power to sexuality is that the overwhelming success, and the limiting consequences, of

the theory of repression hinge on the fact that its mechanisms of power are hidden within its operation. The repressive hypothesis works, which is to say, it has captivated a society of sexualized subjects, under the guise of uncovering fundamental truth and proffering empty liberation. Secrecy, for Foucault, is not just an aspect of power's efficacy but is "indispensable" to its very operation. To put it another way, if power is operating under concealment and seen in an entirely negative, prohibitive way, it is understandable if not reasonable that people could combat one aspect of sexual normalization while remaining ignorant of multiple other ways in which they are enabling and participating as its subject within the same network of power they are attempting to resist. Foucault, rather cynically, suggests that this secrecy, this ignorance, is perhaps just as indispensable to the functioning of the subjugated as the power that dominates them: "would they accept it if they did not see it as a mere limit placed on their desire, leaving a measure of freedom—however slight—intact? Power as a pure limit set on freedom is, at least in our society, the general form of its acceptability" (Foucault 1978, *HS*, p. 86). Power must conceal a substantial part of its mechanisms in order to the gain acceptance of its subjects, but those same subjects must also conceal a substantial part of power from themselves in order to accept the mechanisms of power. However, I do not believe the focus here is on intentional self-deception, but on a positively produced absence of knowledge. What is at stake is a misapprehension of how power how actually works in our society, as well as an underlying faith in the unassailable, rational truth behind our institutions.

One strong implication of Foucault's view is that when power is seen merely as a "pure limit," as long as some small measure of freedom is procured, then all kinds of other constraints will be accepted. This is similar to Foucault's characterization of Kant's concept of Enlightenment as a "contract of rational despotism." When individual subjects view power as purely limiting them rather than actually constituting them through rituals and practices, those very same constitutive mechanisms remain hidden.

What is more, these mechanisms succeed by operating through subjects, and with their own cooperation.

To bring this discussion back to sex, for Foucault, the emergence of sexuality is how society came to be centered on life and organize itself around normalization. What was at issue in the strategies of this great sexual deployment was not the discovery of fundamental knowledge about ourselves through sexuality but "the very production of sexuality." In Foucault's words, "Sexuality must not be sought as a natural given that power holds in check, or as an obscure domain which knowledge tries to gradually uncover. It is the name that can be given to a historical construct…" (Foucault 1978, *HS*, p. 105). And the belief that sex is an "autonomous agency" that power tries to control and make do its bidding is not just a simple misunderstanding; rather, "the idea of 'sex' makes it possible to evade what gives 'power' its power; it enables one to conceive power solely as law and taboo" (Foucault 1978, *HS*, p. 155). If the ability to hide its own mechanisms is indispensable to power's acceptance, with our false belief in sex as a natural given serving as its "most internal element," the deployment of sexuality is one of power's most misunderstood, all-pervasive, and effective mechanisms.

Conclusion: "Outing" Power

"Something smacks of revolt, of promised freedom, of the coming age of a different law, slips easily into this discourse on sexual oppression… Tomorrow sex will be good again" (Foucault 1978, *HS*, p. 7).

Tomorrow, in the post-repressive new dawn of sexual liberation, sex will not be good again, but further normalized. Supposedly transgressive and liberating expressions of sex will be subsumed into the status quo, all but stripped of their subversiveness or played out in a dumb show of deluded provocation and feigned innocence. Tomorrow, sex will not be good again; it will

be more boring.

By concealing its actual manifestations, this deployment of "power-knowledge-pleasure" has gained our consent and convinced us not only that we are sexually repressed but, ironically, that our phony liberation "is in the balance" (Foucault 1978, *HS*, p. 159). So the question becomes how do we not reiterate this repressive fable in slight variations and ever narrowing circles around the norm? How can we resist and dissent from a sexual narrative that not only offers a liberation that does not exist, but beguiles us to participate in our own limitation, of our own accord, in the process? If power is all-pervasive, inextricably intertwined in human social relations, and operates by masking itself, what kind of resistance would be possible, let alone effective?

Foucault's direct answer to the question of resistance is to locate resistance in *bodies and pleasures*: "It is the agency of sex that we must break away from, if we aim—through a tactical reversal of the various mechanisms of sexuality—to counter grips of power with the claims of bodies, pleasures, and knowledges, in their multiplicity and their possibility of resistance. The rallying point for the counterattack against the deployment of sexuality ought not to be sex-desire, but bodies and pleasures" (Foucault 1978, *HS*, p. 157). But what does Foucault mean by *bodies and pleasures*? By his own analysis there is no static material that makes up sex, untainted by history or unfastened from power, to which to refer in a new deployment. It seems that *bodies and pleasures* are what remain after we reject the necessary limitations of the deployment of sexuality.

Judith Butler interprets the phrase *bodies and pleasures* as holding out "the possibility of unmarked bodies," bodies emancipated from the deployment of sexuality (Butler 1999, p. 11). But she is not without critical misgivings for what she views as a problematically hopeful and potentially disappointing rallying cry. For Ladelle McWhorter, the ambiguous yet promising notion of *bodies and pleasures*, from her book by the same name, comes down to the idea that "[b]odies and pleasures have genealogies too" and coming to understand them as genealogies—"as both

concepts and realities [that] play into and help construct the world we live in"—might help us alter that world (McWhorter, *B&S*, p. 136). In other words, if we intend to resist the *dispositif de sexualité'*, "these bodies—these normalized developmental bodies—are the ones that will launch the counterattack" (McWhorter, *B&S*, p. 175).

However, we also must not think that by saying yes to *bodies and pleasures*, one says no to power. The invaluable lesson to take from this text does not reside as much in a new economy of *bodies and pleasures* but in Foucault's critical capacity, in the *outing* of power. Despite what his critics would have you believe, Foucault's genealogy of sexuality is not deficient in allowing for individual agency nor defeatist in terms of resistance simply because he asserts that there is no outside of power and the recognition of what Amy Allen calls the "unavoidable entanglement of power and validity." To the contrary, a kind of resistance is the very function of Foucault's critique, best explicated in his essay, "What is Enlightenment," and put to work here in *The History of Sexuality Vol. 1*. In her discussion of the "nonevent" that was the Foucault/Habermas debate over universalizable notions of rationality and the role of power in subjectivation, Amy Allen, despite her critical misgivings, characterizes Foucault's genealogical project as just such a critique and study of complicated, power-laden social constructions in localized context: "Moreover, although there is a sense in which Foucault does think normalizing, disciplinary power is necessary for creating the modern subject, he also views the aim of his genealogies to be the exposure of what he calls 'the contemporary limits of the necessary,' that is, toward revealing as contingent forms of constraint that are falsely presented as necessary" (Allen 2009, p. 22). In other words, he mounts a critique against *de facto* cultural practices that, despite an absence of valid reason and necessity, operate and function under the presupposition, and hidden veil, of some such rational necessity. And while the creation and practices of the modern subject are never transcendent of power or historical context, for Foucault, there is most definitely an allowance for the possibility of resisting our

perceived "necessary" limitations—in the form of critique.

Foucault is concerned most with acknowledging the entanglement of power processes and social relations, pointing out the strategies and hidden mechanisms of those power processes while at the same time not denying individual agency within them. We've seen in his portrayal of power that it is exercised through individuals and not simply applied to them. As Allen puts it, "Foucault's understanding of individuals as effects of power does not necessitate viewing them as inert, incapable of action, or wholly determined by outside forces" (Allen 2009, p. 22). On the contrary, for Foucault, "they [individuals] are in a position to both submit to and exercise power. They are never the inert or consenting targets of power; they are always also its relays" (Foucault 2003, *SMD*, p. 29). And as I emphasized earlier, the force relations that make up the exercise of power are mobile and unstable, and this allows for the possibility that they can be resisted and changed. Foucault tells us as a general rule of power, and specifically in the explosion of discourse on sexuality, "We must make allowance for the complex and unstable process whereby discourse can be both instrument and an effect of power, but also a hindrance, a stumbling-block, a point of resistance and a starting point for an opposing strategy" (Foucault 1978, *HS*, p. 101).

One might object that this begs the question: "What does this mean we should do concretely to resist?" The answer is perhaps appropriately unsatisfying but potentially more liberating: There are no concrete or specific, positive examples of political action prescribed in Foucault's work; he may not have theoretical recourse to any in a strict sense. Resistance comes in reevaluating the ways in which power has normalized our bodies and revealing our collusion in understanding ourselves through a misunderstanding of that power. The way to resist a reduced and normalized sexuality as natural, given and necessary is to no longer think or act as such; it is what McWhorter characterizes as an ethics or style of *becoming*, and acknowledgment if not celebration of contingency against any static goal or progressive finality that will inevitably prove false—"If we want to oppose normalization,

we should develop disciplinary practices that don't aim at stasis" (McWhorter, *B&P*, p. 193). Contemporary, normalized modes of sexuality perhaps do not lose their significance, efficacy, and performative functions in everyday life against this critique, but they most certainly lose their appeal to the natural and faith in the necessary. What remains in the wake of this critical reckoning is the possibility of emancipated bodies and pleasures and an ethics of contingency. Revealing the operations of power entails a perhaps daunting but utterly liberating freedom.

By diagnosis alone, Foucault takes a look under this mask, showing us something about power's productive capabilities. In so doing, he exposes and undermines this deployment because its successes, and its normative constraints, are contingent on its concealment.

Works Cited

Allen, Amy. (2009) "Discourse, Power, and Subjectivation: The Foucault/Habermas Debate Reconsidered," *Philosophical Forum*, vol. 40, no. 1, pp. 1–28

Butler, Judith. (1999) "Revisiting Bodies and Pleasures," *Theory, Culture & Society*, Vol. 16, No. 2, pp. 11–20

Dreyfus, Hubert L. & Rabinow, Paul. (1983) *Michel Foucault: Beyond Structuralism and Hermeneutics* (Chicago: University of Chicago Press)

Foucault, Michel. (1984) "What is Enlightenment?," in: Rabinow, Paul (Editor), *The Foucault Reader*, pp. 32–50 (New York: Pantheon Books)

Foucault, Michel. (1978) *The History of Sexuality Vol. 1* (New York: Vintage Books)

Foucault, Michel. (2003) *"Society Must Be Defended": Lectures at the College de France, 1975–76*, trans. David Macey (New York: Picador)

Kelley, Sheila. (2003) *The S Factor: Strip Workouts For Every Woman* (New York: Workman Publishing Company, Inc.)

McWhorter, Ladelle. (1999) *Bodies & Pleasures* (Bloomington, Indiana: Indiana University Press)

McWhorter, Ladelle. (2004) "Sex, Race, and Biopower: A Foucauldian Genealogy," *Hypatia*, vol. 19, no. 3, pp. 38–62

Signals of Violence: Contemplating Symbolic Acts of Violence in Postcolonial Discourse

Richard Burton

Introduction

It is a set of acts that begins this inquiry. In the years following the coordinated September 11, 2001 attacks upon the World Trade Center and the Pentagon, postcolonial[1] scholarship has gone to great lengths in accommodating within their work—and critiquing—the vicissitudinous resurgence of discursive practices so powerfully informed by narratives of colonialism, and imperialist systems of domination and control. The further expansion of these discourses, informed by narratives of colonialism and its various forms, transpired in the wake of the Western citizenry's often negative reactions to these events: these unprecedented attacks in New York City and Arlington County bolstered an intense reemergence of nationalism and self-pride, "consolidated by an act of congress: the Patriot Act" (Spivak 2004, 84).

The Patriot Act provided a superficial authorization—signaled by the symbolic possibility of this legislative statute, i.e., an egregious misinterpretation of meaning—misleading the public to respond violently. Thus, a fabricated mode of jurisdiction and justice, and a debased method of protection—one somehow legitimized by racist and discriminative propensities—led to moments of individual and group "patriotism," retaliatory and vengeful in nature. In the immediate wake of the attacks, Anglo-American individuals and groups often prescribed a remedy of violence against "global southerners,[2]" abrogating issues of democracy, civil rights, and universal human rights. A pattern of violence

against global southerners emerged, and the paradigm exemplars, unjustifiably situated at the receiving end of this violence, became, quite literally, *any* person or people perceived to be of Middle Eastern origin/decent. As such, the causal relationship between (a) the Patriot Act and its nascent pattern of nationalistically oriented violence and (b) the persons of Middle Eastern origin having been made paradigm exemplars through this fatally skewed optic, stemmed further theories of innate Anglo-American superiority. Consequences of this racial situatedness, or, perhaps more accurately, the taking of a propertied Western standpoint about various racial groups as normative, fostered a profoundly inimical terrain in which this pattern of violence could be sustained. At the forefront of our minds are these obvious signals of violence: violent acts performed by clearly definable agents, reified in the context of hate crimes.

Nearly ten years after the events of September 11th, 2001, these narratives of racialized violence remain by virtue of testimonies, solemnly uttered by those willing. However, empirically verifiable moments of violence aside, delusions of racial—and often national—superiority insulate themselves from refutation. It is the Žižekian mode of inquiry into violence that attends to what he notes as the "contours" of violence, or, more accurately, the above delusions concerning race. A manner of critique and elaboration in this style moves away from the most visible portions of violence—for our purposes, racially motivated instances—which "shock" or "disturb." Instead, such a departing mode of critique concerns the deconstruction and exposition of violence that is ideological, and this ideology must be exposed as a contingent matter, existing as a continuation of colonial discourse, a consequence of the "smooth functioning" of our economic and political systems (Žižek 2008, 2).

Thus, this essay is threefold. In Section (1), I begin with a preliminary sketch of decolonization, viz., the autonomous act of revealing and dismantling colonialist power, including the hidden traces of those institutional and cultural models that sustain a pervasive, colonialist influence even after the withdrawal of a

colonial power from a colonized population *qua* the achievement of political self-determination. Arguably, the implausibility of a normative account of a decolonizing method—a concept so fundamental to the factical possibility of a post-colony—prevails. Bearing in mind how history has heavily emphasized the gravity of the colonial encounter, situating it within the dialectic of domination, of repression, of exploitation, etc., an analysis of the status quo illuminates the possibility of total decolonization as a point of extreme contention among scholars. Due to the view that colonial practices, discursive or material, remain a viable institution, the postcolonial enterprise is comprised of critics and scholars who utilize their energies to illuminate how the underpinnings of the decolonizing apparatus poses a challenging ethical problem.

In Section (2), with the model of September 11th, 2001 as a specific moment of concrete, observable violence, I turn to a demonstrative example of contoured, symbolic violence, one motivated by Western, postmodern media's blatant lack of sensitivity and inherent presuppositions about divergent cultures. I draw on the state of emergency Haiti found itself in after being rocked by a massive earthquake in early 2010 vis-à-vis its depiction by Western media outlets. With this framework in place, I will be able to assert my argument that CNN news anchor, Anderson Cooper, engenders and confronts us with a variant of unconditional, symbolic violence throughout a definable act of heroic intervention. In defending the human rights of "the Other[3]," Cooper imposes his own version onto the Other, the operations of which are predicated on the principles of Western normativity. Within this ideological space, the privileging of this certain (heterosexual, Western, Anglo-Saxon) culture spills over into the realm of this Other (the most pertinent constituents of which are self-constructive autonomy, rationality, and subjectivity) through their discursive construction of the Other's identity as something contingent on Western assumptions and stereotypes. Thus, the premises of this benevolent, lionhearted gesture force a discursively oppressive and exploitative conclusion. This demonstrative account exemplifies the relationship of the West and the West's

other as an operational mode of colonial practices, modeled within a wide range of institutions, maintained here, particularly, by the Western media.

Finally, in section (3), with this framework in place, I will begin to shape a hypothesis predicated on the future of postcolonial studies and possible ways it might continue to accommodate the resurgence of colonial discourse. I conclude that the provision of a bolstered nexus between the public and private sphere, or, more accurately, between the highly specialized academy in which these studies reside and Western society, might have a more edifying, enlightening effect.

Section 1.
Decolonization's Critical Enterprise

"We live in a world where there is more and more information, and less and less meaning" (Baudrillard 1994, 79)

In the opening of his essay *Freedom's Basis in the Indeterminate*, Homi K. Bhabha offers a poignant account of the ways in which postcolonial criticism bears witness to the dissimilar and unbalanced forces of cultural representation. A primary goal of the postcolonial intellectual is combating the residuum of colonialism within cultures. In other words, postcolonial scholarship does not *only* attend the salvaging of past worlds, but also attempts to generate a method of moving beyond a historically delimited sense of postcoloniality; postcolonial scholars want to explicate that colonialism's fundamental anchorage is not limited *only* in the topic of history due to the logic of colonialism going un-accounted for in un-recognized forms.

The most flagrant evidence of this logic's ramifications are expressed in the struggle for sociopolitical recognition, legally sanctioned authority, and ontological self-determination by minorities ordered within the geopolitical divisions of East and West, North and South (Bhabha 1994: 171, 173). Discourses taking place within the "modern world order" reveal postcolonial perspectives that are a result of postcolonial or anticolonial testimonies

(Bhabha 1992, 46). Thus in bending Bhabha to our purposes we can explore my argument of the posturing of postcolonial scholarship as it stands–Third World intellectuals in First World academe who are congenitally connected to this linear cultural history[4]—informs a discursive strategy that operates against the normalizing and naturalizing tendencies of modern Western ideologies. Simultaneously, this possibility for intercession on the postcolonial intellectual's part signals an instance supportive of a totalized, determinative postcolonial subject; a corroboration that registers on all levels.

Whether through reading, writing, teaching, criticism, or activism, a deliberate intervention in hegemonic discursive and material practices is staged by champions of postcolonialism and anti-colonialism alike. However, this is not to suggest that scholars of postcolonial studies are to be crudely qualified via geographical origin; doing so renders the phenomena of this scholarship in a singular dimension. Moreover, the emergence of postcolonial studies within the 1980s was certainly multi-dimensional, informed and structured by proliferations within cultural/area studies; scholarship which was further disseminated throughout the humanities and social sciences.

Historian Arif Dirlik concisely frames a necessarily fluid identity of postcolonial intellectualism as a "perspective [that] represents an attempt to regroup intellectuals of uncertain location under the banner of postcolonial discourse. Intellectuals in the flesh may produce the themes that constitute postcolonial discourse, but it is participation in the discourse that defines them as postcolonial intellectuals" (Arif Dirlik 332). Through the use of temporal words and phrases—"in the flesh," for instance–Dirlik draws attention to how the term postcolonial no longer references *only a* general time or space, that is one, historically speaking, *after* colonialism. Such an interpretation reveals that "postcolonial" thinking has ceased to be a merely historical category. The postcolonial perspective departs from this historically delimited sense, one interested solely in erstwhile colonial territories that have effectively been decolonized to "postcolonial" states–states

capable of stimulating and sustaining independent development. By pinpointing the more complex cultural and political boundaries that still define the First World and Third World's relationship, this perspective attempts a forced recognition of a historically exclusive notion of postcolonialism's entailments as insufficient (Bhabha 1994: 173). If the West is to be understood as ideologically driven, its manifestations are, then, subject to the critique of postcolonial intellectuals..

In an attempt to transform the world created by Western political and economic hegemony, the enterprise of postcolonial intellectuals illuminates the apparatus of decolonization's underpinnings, the conditions of which are continually affected by the existence of colonial and imperial antagonisms. For postcolonial intellectuals, the articulation of a sufficient deconstructive method—apropos the *total* decolonization of colonial realities—remains frustratingly incomplete.

During the last half of the 20th century, the pervasive national, cultural, and political forms of globalized, Western influence have further inhibited progress. Numerous consequences of these socio-political narratives, envisioned through the hegemonic imperative of the West, forcefully diffuse re-figured and re-formulated forms of knowledge and social identities, authored and authorized by the most salient inequities in postcoloniality: the residuum of colonialism and western domination as typified by neocolonialism. These neocolonialist currents elicit new modes and forms of colonial ideology, within contemporary postcolonial cultures.

Through the continuous and multifarious narratives of cultural oppression, subjugation and diasporic displacement, the inherent complexities of a decolonizing concept sustain a malignant paradox that curtails the formulate emergence of a *total* decolonizing process. Arguably, Westernized global discourse prolongs the conundrum of decolonization, or, more accurately, protracts the impossibility for formerly colonized nations to truly reveal and dismantle colonialist power in all its forms, even after the withdrawing of a colonizing power

Addressing the discursive difference of people from divergent cultures and those clearly rooted in a Western context, Bhabha writes, "Culture-as-sign articulates the in-between moment when the rule of language as semiotic system–linguistic difference, the arbitrariness of the sign–turns into a struggle for the historical and ethical *right to signify*" (Bhabha 1992, 49). Primarily, signifiers of homogeneity and hegemony coerce the ethics of signification. In the interest of socio-economic ascendancy, colonizing powers establish these pursuits as normative actions beneficial to both colonizer and colonized. Through the naturalization of these economic aims, the colonizing body deliberately veils the perverse reality of this social relationship's constitution. More accurately, through acts of total self-interest, the colonizing force negligently fails in an egalitarian representation of divergent, native populations, consequently forcing a dominant/subordinate division.

According to Bhabha, situating any engagement with culture, primarily with that of the Other—whether this "Other" is the Orient (Said 1978), or the Subaltern (Spivak 1988)—is a strategy for social survival. Culture as a strategy of survival functions in both a *transnational* and *translational* mode. The transnational element of this dyad is rooted in and facilitates the specific histories of displaced cultures within the context of contemporary colonial discourse: "…the 'middle passage' of slavery and indenture; in the 'voyage out' of the colonialist's civilizing mission; in the fraught accommodation of postwar 'third world' migration to the West; or in the traffic of economic and political refugees within and outside the Third World" (Bhabha 1992, 47). The latter dimension is *translational* in the sense that it interrogates these histories of cultural displacement by leading the question of *how* culture signifies, or what is signified by *culture.*

Scholars in postcolonial studies have tended to address the subsequent answer to this question as one situated by a hegemonic conceptualization of cultural Otherness, viz., domination through the interpellation of colonized subjects by Western, imperial discourse(s). Regarding this vexing question of semiology[5] further, Bhabha frames the parameters of cultural identity, particu-

larly that of the individual, as a sign in which "negative ontologies" may be inscribed (Bhabha 1992, 46). If we are faced with a notion of structural violence attached to a Western, normalized discourse as opposed to intentional modes of violence as an expression of a collectivity of people, then violence is instead inscribed in the dialectic of identity and Otherness.

Building on this formulation, Bhabha's putative critique is consistently informed through study of the discursive dichotomies of West/East and North/South. On the basis of these accounts, Western assumptions about non-Western, non-European taxonomies are continually bolstered by numerous predilections such as "spiritual values, cultural assumptions, social discrimination, racial prejudices and humanistic values" (Ashcroft, et al, 107-108).

As I have noted above, dealings with violence in postcolonial discourse, self-determination in symbolic signification, which is a central activity of subjectivity, is denied to colonized peoples through the continuation of a neocolonial discourse. Arguably, at one end of this symbolic spectrum, the violence immediately following the attacks on September 11th, 2001 can be situated. Violence here was expressed through individual, as well as collective acts of violence. Those victimized were explicitly denied self-determinative signification, the chief source of which was the forced construction of racially stereotyped identities, or, more appropriately, "negative ontologies." At the other end of the spectrum, and the crux of this essay, are the subtler, contoured expressions of violence.

Section 2.
Media, Postmodernity and the Ethics of Signification

After being struck by the most severe earthquake in over 200 years, the Republic of Haiti had gone from a state of financial despair, "…an economic wreck, balancing precariously on the edge of calamity," to one of near physical non-existence[6] (Farnsworth 2010). Even with the full ramifications and urgency of the

Haiti crisis becoming apparent, the arrival of humanitarian aid was minimal at first, decelerated by a number of circumstances. Within hours of the earthquake, Cuban doctors, Chinese search and rescue teams and Venezuelan medical professionals were the first respondents to arrive in Haiti. Yet, when the United States military took control of the Port-au-Prince Airport, they prioritized an immediate landing of U.S. troops instead of a mobilization of humanitarian resources. Arguably, this hesitancy by the United States was due to wariness about the possibility of violent acts being committed by the Haitian people. Yet, finally, with a now nearly razed capital city and an untold number of causalities, the United States, France, Japan, the European Union and the United Nations reached the distressed country to begin providing aid.

The enunciation of such an irrational fear—the lure, and simultaneous threat of people other to the West—discursively maps, through various modalities, a counter-narrative onto the physical space of Haiti. In effect, the country became presented as an *objet d'art*–a physical space fulfilling an independent and primarily aesthetic function. Emptied of meaning, this re-constitution of Haiti as a domain, now residing solely within the Western imagination, allowed for the coercive inscribing of identity and Otherness. The arrival of this international alliance instantaneously signaled the entry of territorial ambitions of global media technologies. In other words, news media organizations had arrived in Haiti, transporting and transferring a dynamic exchange of culturally imperialist ideologies.

To better understand the imperialist ramifications of mass media from a postcolonial perspective, or its considerable role in postmodern discourse, Edward Said provides a telling critique:

> One aspect of the electronic, postmodern world is that there has been a reinforcement of the stereotypes by which the Orient is viewed. Television, the films, and all the media's resources have forced information into more and more standardized molds. So far as the Orient is concerned, stan-

> dardization and cultural stereotyping have intensified the hold of the nineteenth-century academic and imaginative demonology of "the mysterious Orient" (Said 1978, 4).

Though situated by the placeholder "Orient," Said's analysis illuminates the "reinforcement" of postmodern media's tendency to pigeonhole any cultural identity/information that resides within any body politic(s) of non-European descent who have been determined by, and fixed outside of a naturalized, Western citizenry. For Said, the Orient is, "not an inert fact of nature. It is not merely *there*, just as the Occident itself is not just *there* either" (Said 1978, 4). Thus, the Orient exists merely as a discursive realm of colonial imagination within a larger catalogue of innumerable, disparate postcolonial narratives. In effect, critical revisions of postcolonial perspectives are inherently fluid, and are appropriately re-formulated around the issues of cultural dissimilarity (distinct colonial/anti-colonial narratives), persecution (colonial/imperial oppression), and most importantly for the purposes of this essay, the "ambivalent moments within the 'rationalizations' of modernity" (Bhabha 1992, 46). These moments of ambivalence can be characterized as a complex mix of attraction and repulsion between the colonizer and colonized, and for our purposes, the Western audience and the physical presence in Haiti–bodies. If we accept this currency of ambivalence (how Westerners are simultaneously attracted to the plight of non-Westerners and threatened by non-Westerners due to there radical alterity–the fact they are Other to the Western self) its function announces itself as a method for Western audiences, so absorbed by ideological imperatives, to explain or logically justify these moments. The discourse of colonialism is pervaded by images of extreme violence, rape, etc., which are continually grounded by the West's persistent fantasizing of the other.

Befogged by ideology and notions of divergent cultures as generated by Western media, the legitimacy of the Other's identity emerges as a forced universality, contingent on Western media viewer's limited perspective. Unable to escape the symbolic

fictions expressed by the Western media's framing of nonnative social conditions, the possibility for the non-Western Other to enlighten this limited perspective through symbolic signification in un-inhibited self-determination is effectively silenced. After the initial earthquake struck Haiti, CNN's anchor Anderson Cooper, along with a film crew, were dispatched to cover the aftermath. Six days after the quake in the municipality of Port-au-Prince, Cooper and his crew witnessed and filmed the looting of a small shop, from which a group of young Haitian men had pilfered a few candles in order to sell them on the street. In an attempt to disperse the crowd, Port-au-Prince authorities began firing warning shots skyward. Hearing the shots overhead, a few agitated looters took cover on the shop roof and began propelling down chunks of rubble from their makeshift vantage. A young boy was struck in the head by a falling piece of concrete, lacerating his face and knocking him to the ground. In a cinematic moment of hyperreality, Anderson Cooper immediately responded, sprinting down the street to the fallen boy's aid. Later, in a CNN news report, Cooper stated, "If [the boy] stayed there, he might have gotten killed." Followed by two other men, Cooper drags the young boy away from the scene and further down the street, toward an improvised barrier. When the boy is finally allowed to act, he immediately recoils from Cooper and wipes the blood away from his face. Cooper, frantically scanning his surroundings, repeatedly claps the boys shoulder and repeats, "It's okay… it's okay… it's okay." In a later commentary, Cooper states how the boy was "obviously stunned" and "[couldn't] walk," yet, in contrast, the video reveals the injured boy's desire to be anywhere but under this guise of safety; here, the injured boy's display of autonomy and agency fundamentally disrupts and challenges the self-conscious account delivered by Cooper. In a final expression of stable, "ethical" agency, Cooper lifts the boy off the ground, and cradling him in his arms makes the final dash for the barrier. At the barrier, the boy is removed from Cooper's custody and led away. It is exactly here that "culture-as-sign," that of the injured boy, becomes *the* moment of violent struggle for access to both

history and the right to signification.[7] The crucial point here is that the film of Cooper's interference to save the boy privileges a discourse of Western superiority and universality that masks the authenticity of the Other and his culture; the Other's freedom to choose is revoked, or, more accurately, so perverted by the operations of Western domination, he is ripped out of his existence by the violent over-determination of Western society. This episode signifies Cooper as Western "savior," eliminating the boy's ability, and, more broadly, of the Haitian people, to help themselves; thus, theories of Western superiority stem further from this depiction of the West's Other.

Seen through the perspective of postcolonial theorists who study colonialism in all its diverse expression, the rationalizing apparatus, signaled by the modern Western audience, incite myriad feelings towards a person or object, i.e., sympathy, fear, sadness, pride, and in the instances of Haiti and postmodern media, as a universally signifying apparatus; this level of emotional disturbance is thus leveled at the symbolic interaction between an ambassador of Western discourse (a journalist) and a Haitian boy– the West's other. Emphasizing this disabling moment of Western rationality, I assert that the proclaimed heroism of Anderson Cooper (an ambassador of a benevolent, Western centre) and the aid he provides, promotes a discursive mechanism of power which violently constructs the colonial Other through the binary logic, or, for our purposes, the rigid ideological hierarchy of colonizer and colonized, effectively denying the colonial subject symbolic signification; this denial is privileged through rationalizing Cooper as a superior subject, even though he discursively posits, through his actions, a kind of putative, naturalized quality to his current situation. Consequently, a symbolic discourse results from contemporary configurations of Western colonial practices in the midst of their supposed disintegration.

Section 3.
Violence and the Symbolic: Making sense of "silent" interaction

In *Violence*, philosopher Slavoj Žižek reflects on the paradox of subjective, visible violence, and acts of ideological violence within the discourse of postmodernity.

> At the forefront of our minds, signals of violence are acts of crime and terror, civil unrest, international conflict. But we should learn to step back, to disentangle ourselves from the fascinating lure of this directly visible "subjective" violence, violence performed by a clearly identifiable agent […] we need to perceive the contours of the background which generates such outbursts. A step back enables us to identify a violence that sustains our very efforts to fight violence and to promote tolerance (Žižek 2008, 1).

Here, Žižek is pointing to the fact that "signals of violence" in any traditional sense–night vision footage of shelling in Iraq, race riots in the streets of Los Angeles, or police brutality in Mexico City–in fact entangle one's sensibilities by luring a subject to more crystalline, often relatable junctures in violent acts. In other words, the spectacle of immediately definable acts of violence allow us to comprehend the contours of an immediate situation, and inscribe stereotypical levels of understanding and cultural integrity onto "clearly identifiable agents" of violence. Subjective modes of violence are "just the most visible portion of a triumvirate that also includes two objective kinds of violence." For Žižek, a fundamental form of violence, which he elaborates as an objective mode, elides this specific mechanism of codification–our linguistic account of the visible portions of violence–as it is "precisely the violence inherent to [a] normal state of things" (Žižek 2008: 1, 2).

In Žižek's account of subjective and objective modes of violence, he draws close attention to the latter formulation, a notion of violence produced and governed through ideological

discursive regimes of power. Within the "invisible" realm of objective violence, a discourse endemic within the postmodern world order, exist two strata. For our purpose, the "symbolic" level of violence embodied in language and its forms–language, communication, rhetoric, image, media, et al–demands careful consideration. Žižek claims that language itself involves unconditional violence since "when we perceive something as an act of violence, we measure it by a presupposed standard of what the 'normal' non-violent standard is–and the highest form of violence is the imposition of this standard with reference to which some events appear as 'violent.'" Žižek draws on Jacques Lacan's work on the ethics of psychoanalysis by stating "it is language itself which pushes our desire beyond proper limits, transforming it into a 'desire that contains the infinite'" (Žižek 2008: 64-65). Desire transcends "proper" limits and transforms into an un-containable "infinite," implying a violent disturbance in language's (and its forms) interminable extension between signifier and signifier. Secondly, Žižek argues for the existence of systemic violence, "or often the catastrophic consequences of the smooth functioning of our economic and political systems" (Žižek 2008, 2).

This expansion of subjective and objective realities of violence provokes the anecdote concerning the discursive relationship between Haiti (embodied in the Western "other," viz., the compromised Haitian boy) and Western humanitarian/media outlets (constituted by the benevolent quester, Anderson Cooper). The event that took place on January 16th—the Anglo-American male "saving" a young Haitian boy from a violent upheaval—was a symbolic performance. However, this appearance of heroism and benevolence dissimulates the brutal "reality" of the street scene and overdetermines any political or humanitarian consideration. This is a moment where "information devours its own content. It devours communication and the social" (Baudrillard 1994, 80). The Western audience, in receiving Cooper and his film crew's performative information, glean an interpretation cast from a mold of Western hegemonic ideology in which socially communicable information can be of no importance. In other words,

any scene of cultural difference initiates in the Western audience a "devouring" apparatus of universal signification in which these moments of ambivalence and uncertainty become diffused through the inscription of "universal," or, more accurately, naturalized meaning in an attempt to "normalize" the other by way of complete Westernization.

The narrative in question–the beaten, bloodied boy–incites a myriad of emotionally driven qualifiers: helpless, harmless, vulnerable, unsafe, and violent. Žižek insists these linguistic impositions form the "highest form of violence," effectively perpetuating ideological misreadings and cultural misrepresentations through the act of silencing and signifying for the other. For philosopher Jean Baudrillard, these qualifiers, spoken by the spectral voice of the Western audience, "[have] nothing to do with signification. It is something else, an operational mode of another order, outside meaning and of the circulation of meaning strictly speaking" (Baudrillard 1994, 79). If ideologically influenced mis-significations of cultural meaning by the West continue their existence, then the discourse of colonial oppression will remain unshaken. Furthermore, the transferring of meaning, and arguably information, into the realm of the Western imagination only serves to strengthen Western misunderstandings and stereotypes by rendering it even less susceptible to analysis. Where a postcolonial semiology is considered, the dependency of signs upon other signs is the method by which they derive their meaning from their relation to other words. However, if the signs in this order are fixed outside of meaning and the very circulation of meaning, then an "in-between" moment is signaled, one in which the semiotic rules of language and its forms become a "moment of struggle." In other words, the subject considered here becomes re-typified as surrogate victim. Such circumstances bring to mind the types of violence that break out spontaneously in countries convulsed by crisis. However, it is not enough to say that this individual is exposed to an explicit act of collective, public violence by way of the rioting crowd, or private violence as a result of the causal relationship between himself and Anderson Cooper; any display of

violence verifiable through observation remains secondary to the symbolic ramification which contextualize this particular account. Contrarily, the mechanism of symbolic violence—and arguably the endemic lack of recognition in regards to its existence—can be delineated as an aberration in the colonial legacy in any postcolonial society. Whether a consideration of a unique moment in Haiti, the authenticity of violence as *understood* in occupied Palestine, or the instability, corruption and authoritarianism still strangling "postcolonial" Africa, the more transparent, subjectively definable instance of observable violence remains the instrument by which degrees of brutality, cruelty, oppression, etc., are evaluated.

If we consider the dynamic symbolic exchange between the young Haitian boy and Anderson Cooper and the ways in which symbolic violence is valorized within the postcolonial discourse, the semiotic struggle is elucidated further. The question that I am interested in is that of Mr. Cooper's ethical agency and privileged existence as precisely a relation that reduces the young boy to silence by cluttering his "speech" with signs of undoing. Within the brief time span of the news footage, a performative act of violent symbiosis is carried out against the boy's agency, compromising the integrity of his ontology.

In this moment, cultural imperialism's "image as the establisher of the good society" emerges as an obvious position of complete and utter dominance: Anderson Cooper literally speaks for the boy, robbing him of voice and his totality of symbolic possibility (Spivak 1988, 299). Reiterating Žižek, symbolic acts of violence within ideological discourse are *unconditionally* violent. Here, the proper limits of desiring are pushed beyond the threshold, producing a rupture within the "'wall of language' which forever separates me from the abyss of another subject [and] is simultaneously that which opens up and sustains this abyss—the very obstacle that separates [the subject] from the Beyond is what creates its mirage" (Žižek 2008, 73)[8]. Derived from Lacanian psychoanalysis, the maxim "wall of language" more accurately represents the totality of factical possibilities which spur the subject to authentically signify in the world. Derived from

a disturbance in the "wall of language," the construction of the boy's identity is discursively determined by his ideological union with "rational" and "primal" Anderson Cooper.

Rather than an inquiry of how he might go about helping in a moment of such confusion, e.g., what plan of action would be appropriate/suitable for the circumstances, a role he could sufficiently perform, etc., Mr. Cooper, pandering instead to Western ideologies of service and aid, espouses the "helpless" boy as an object of protection from his own kind. Cooper is a Western reporter, and the desire to intervene is a (neo)colonial desire, even in this beneficiary mode. Consequently, he enacts an obligation to rule over another, even in the symbolic. With this patriarchal strategy, possibility of free choice by the subject becomes totalized within the more powerful discourse of the postcolonial continuum. Philosopher Judith Butler writes that the body:

> Implies mortality, vulnerability, agency: the skin and the flesh expose us to the gaze of other, but also to touch, and to violence, and bodies put us at risk of becoming the agency and instrument of all these as well […] the very bodies for which we struggle are not quite ever only our own. The body has its invariably public dimension (Butler 2004, 26).

All this is not to merely say that a corporeal subject can be compromised or negated at any given moment. Rather, this citation speaks to the manipulative possibilities that a body is "capable" of in the domain of public discourse. The body is most certainly *that*, a public *thing*, and the body of the bloodied Haitian boy is generated by the multitude of Western practices, becoming merely an empty signifier in which Western misunderstanding(s) of the "Orient," non-European "other," and subaltern can be freely inscribed, viz., racially charged inscriptions of identity and Otherness. The West's aspiration to represent non-Westerners in Western terms articulates what is wrong in postcolonial realities. The locus of critical inquiry is *not* that a North American journalist aided someone in obvious mortal peril, but instead it resides within the

issue of Western ideologies' continually violent construction of the "other" as enclosed within the logical construct of a colonial/imperial binary and a constant obfuscation of self-determinant modes of "cultures of difference."

However, the "events" in Haiti prove that the philosophical imperative of critical inquiry and examination are increasingly relevant in our "postcolonial," "postmodern" society. In regards to conflicts within divergent cultures, these moments of symbolic, postcolonial violence naturalize a certain feeling of Western superiority through their repetition within the dominant discourse of Western imperial supremacy. Although this paper does not offer a solution for a totalized vision of decolonization within colonial realties, it does begin to formulate a hypothesis critically aware of the effects of 21st century postcolonial scholarship, media, news sources, and political rhetoric. More accurately, I offer the following beginnings of hypothesis for the way in which postcolonial scholars might accommodate modern colonial realities in their discourse: In order to begin analyzing representations of divergent cultures,—whilst eliding the white man's burden and Otherness—with the hope of decolonizing the residual logic of colonial discourse within postcolonial realities, an enormous shift in global understanding of perpetual ideology must be situated at the fore of scholarship, and our critical consciousness. This entails a radical deconstruction of those hidden aspects of colonialism that so fervently maintain the implausibility of a total decolonizing method.

Works Cited

Ashcroft, Bill, Gareth Griffiths, and Helen Tiffin. *Post-colonial Studies: the Key Concepts*. London: Routledge, 2000. Print.

Baudrillard, Jean. *Simulacra and Simulation*. Ann Arbor: University of Michigan, 1994. Print.

Bhabha, Homi K. "Freedom's Basis in the Indeterminate." October 61 (1991): 46–57. Print.

_____. *The Location of Culture*. London: Routledge, 2004. Print.

Butler, Judith. *Precarious Life: the Powers of Mourning and Violence*. London: Verso, 2004. Print.

Cooper, Anderson. "In the Midst of Looting Chaos." Video Blog post. Anderson Cooper 360. Cnn.com, 18 Jan. 2010. Web. Feb. 2010. <http://ac360.blogs.cnn.com/2010/01/18/anderson-in-the-midst-of-looting-chaos/>.

Dirlik, Arif. "The Postcolonial Aura: Third World Criticism in the Age of Global Capitalism." Winter 20 (1994): 328-356. Print.

Farnsworth, Eric. "Reconstruction Will Be the Real Challenge." National Journal Online (2010). Americas Society Website. Council of the Americas, 19 Jan. 2010. Web. Feb. 2010. <http://www.americas-society.org/article.php?id=22115>.

Said, Edward W. *Orientalism*. New York: Pantheon, 1978. Print.

Spivak, Gayatri Chakravorty. " Can the Subaltern Speak?" in Marxism and the Interpretation of Culture. Eds. Cary Nelson and Lawrence Grossberg. Urbana, IL: University of Illinois Press, 1988: 271-313.

Žižek, Slavoj. *Violence: Six Sideways Reflections*. New York: Picador, 2008. Print.

Notes

1. The term(s) post-colonialism/postcolonialism is a site of disciplinary and interpretive contestation. A working definition of postcolonialism in its historical sense, and theoretical sense, is, "Post-colonialism (or often postcolonialism) deals with the effects of colonization on cultures and societies. As originally used by historians after the Second World War in terms such as the post-colonial state, 'post-colonial' had a clearly chronological meaning, designating the post-independence period. However, from the late 1970s the term has been used by literary critics to discuss the various cultural effects of colonization […] An equally fundamental constraint is attention to precise location. Every colonial encounter or 'contact zone' is different, and each 'post-colonial'occasion needs,against these general background principles,to be precisely located and analysed for its specific interplay. A vigorous debate has revolved around the potentially homogenizing effect of the term 'post-colonial' (Hodge and Mishra 1990; Chrisman and Williams 1993). The effect of describing the colonial experience of a great range of cultures by this term, it is argued, is to elide the differences between them. However, there is no inherent or inevitable reason for this to occur. The materiality and locality of various kinds of post-colonial experience are precisely what provide the richest potential for post-colonial studies, and they enable the specific analysis of the various effects of colonial discourse" (Ashcroft, et al., 2000, 155–156).

2. A socio-economic and political divide between countries demarcated by the "wealthier" Northern hemisphere and "poorer" Southern hemisphere.
3. In Western philosophy, and critical theory, "Other" connotes myriad definitions. Fundamentally, the Other is anyone who is separate from the subject and one's self. The existence of others is crucial in defining the individual subject and, for our purposes, the employment of Other in this essay will qualify what I find to be a dichotomous relationship between the "naturalness" and "primacy" of Western cultural institutions and those existing outside of Western culture and the colonizing world view, in particular the native population of Haiti.
4. My primary concern is with contemporary intellectuals based in the continental United States (Homi Bhabha, Giyatri Spivak, Edward Said et al.) who face the inherent advantages and disadvantages set forth by Western academe, but are culturally and historically located within a geopolitically oriented, postcolonial discourse.
5. Ferdinand de Saussure isolates language as an autonomous object of study in his posthumous work *Course in General Linguistics.*
6. However, Haiti going from a state of economic, financial despair, to a state of utter ruin in the wake of natural disaster should *not* depict these states as "naturalized." Haiti is a country deeply rooted in revolution. In "Toward a Second Haitian Revolution," a recent article in Harper's magazine, Steven Stoll writes, "Haiti's agony began with its success. As a plantation colony, it was the richest in the hemisphere. Beginning in the seventeenth century, the planters of French St. Domingue—the island's western third—amassed land and labor, producing coffee, sugar and cotton with slaves they brought over from West Africa. The plantation regime crumbled in 1791, when the slaves stunned and terrifies the Americal and European elite by enacting the most radical principles of the Enlightenment. They grabbed whips and hot irons out of the hands of their overseers, hanged their colonial overlords, fought off a british invasion, and defeated Napolean's army of occupation before declaring independence in 1804. No subjugated people had ever so upended the social order, and no one who had profited from that order ever forgave them" (Stoll, *Harper's Magazine,* 2010, 7)
7. All video footage can be found at the following link: <http://ac360.blogs.cnn.com/2010/01/18/anderson-in-the-midst-of-looting-chaos/>.
8. Photo taken from Mr. Coopers blog: <http://ac360.blogs.cnn.com/2010/01/18/anderson-in-the-midst-of-looting-chaos/>.

Modularity and the Cognitive Architecture of Mind: Fodorian Modules and The Niche Construction Theory of Sterelny

Kyle Arneson

Introduction

Understanding the neural architecture of the mind has brought about numerous competing models of cognition within the philosophy of science, two of which will be considered here. The first, by Jerry Fodor, contains a lower-level input system of modules functionally individuated from higher-level cognitive domains and the second, by Kim Sterelny, relies on a wider range of central processing. I will argue that both are computationally intractable and fail to reasonably satisfy the frame problem. The paper will proceed with an articulation of both cognitive models, Fodor representing the nativist view and Sterelny the non-nativist, divided by the poverty of the stimulus argument acting as the fulcrum to the debate. The notion of modularity plays a role of considerable philosophical import in numerous ongoing debates, appearing as a contributing factor in theory-neutral observation styles in the field of epistemology, arguments for scientific realism and eliminativism in philosophy of science, as well as countless instances of reliance on and reference to philosophy of language (Robbins, 2009). Getting clear about the tenability of cognitive modules might yield philosophical advances in any one of these platforms. A look in turn at massively modular, then moderately modular, and finally non-modular systems of cognitive architecture, with criticisms of each, allows for the advancement of potentially tractable alternatives.

§1. Massive Modularity and Faculty Organization: Vertical/Horizontal

The Massive Modularity Hypothesis is a theoretical mental architecture that differs significantly from a Fodorian complex of input system modules given its greater operative division. Because neither Sterelny nor Fodor regard Massive Modularity to be a tenable schema of cognitive organization it will be mentioned and briefly sketched only to be set aside. The Massive Modularity Hypothesis is primarily useful as a point from which to delineate Fodor's thesis before turning to Sterelny's refutation, the arc of which requires a dismantling of any massively modular cognitive theory. Robust and vigorously defended theories of Massive Modularity, like those put forth by Peter Carruthers, consider the mind to be comprehensively segmental[1] with respect to modular functionality including areas of higher-level cognitive domain. Modules under this theory depart from Fodorian versions by being isolable, function-particular processing systems that are mostly domain specific. They are unencumbered by willful action but associated with specific neural networks whose internal operations are inaccessible to the remainder of the cognitive path. Sterelny discredits any cognitive account on this level for the simple reason of evolutionary incongruity: While the language portion of the brain presently appears as inexplicable without a modular schema (as will be explored below), to consider every mode of cognition as likewise constituted would result in numerous, unaccountable brute operations upon which millennia of evolution would have held no sway. To have modularity of the mind at this level would entail dedicated actions and circuitry forged not by the slow advancement of human development but rather by an unaccountable, albeit clearly designated function (i.e. a module for speech, a module for cooking, a module for playing chess, etc.) before such actions were developed, honed, or even conceived of. Fodor, while holding to the tenability of input system modularity, rejects the plausibility of massive or central modularity given their Quinean nature and resulting need

for entire belief system confirmation (i.e. multi-layered verification and cross-referencing definitionally unavailable to modules). More will be said about this below. The features of the Massive Modularity Hypothesis shared by Fodorian models include dissociability, domain specificity, mandatoriness and central inaccessibility. Notably missing from this list, however, is information encapsulation, the principal beam of Fodor's framework. These features will be covered in turn.

The horizontal/vertical format of mental composition refers to a functional rather than spatial model of cognitive relationships. A horizontal faculty model is functionally arranged and distinguished from a vertical system by its inter-domain content sharing capability. Cognitive processes under this format make shared neuronal content accessible to and by any mental process in the strong sense (what Fodor regards as "thoroughly horizontal") or by most mental processes in the conventional sense (Fodor 1983, p. 13). Content under this model is inherently non-domain specific as higher-order faculties (memory, judgment, belief, etc.) avail themselves to the content freely. Conversely, vertical faculties possess the notable characteristics of domain specificity and computational autonomy and are thought to be genetically determined as well as identifiable given the distinct neural structures needed to perform these specified tasks. Franz Joseph Gall, from whom Fodor derives much initial inspiration[2], posited early and memorable versions of the horizontal/vertical model-contrast and from Gall we derive Fodor's refined vertical model of input system modularity.

§2. Fodorian Modularity: Information Encapsulation and Domain Specificity

Fodor's model represents a form of psychological nativism wherein certain innate cognitive functions are computationally autonomous and therefore unaffected by any encroachment of preliminary mental processes or subsequent cognitive domains. Broadly, his thesis is in part a reaction to associationism, containing

elements of or refinements to Chomsky's Neocartesianism and other similarly veined vertical accounts of faculty psychology (that is, accounts which differentiate cognitive functions based on encapsulation and neural structure as opposed to broad interaction between such domains as judgment, memory and so forth). Mental processes, according to Fodor, rather than being operationally necessitated by successive higher-level processes are instead operationally individuated by domain specific functions with impermeable information encapsulation. From this we can derive and outline a subset of distinct features within Fodorian modularity including the aforementioned domain specificity and information encapsulation, as well as mandatory operation with limited medial accessibility, rapid processing (given a fixed neural architecture), and specific breakdown patterns with identifiable ontogenetic rhythms and sequences. A qualifying modular system of cognition, according to Fodor, contains at least a majority of these components, most importantly of which is information encapsulation. Encapsulation, being what truly distinguishes modules from other cognitive domains, refers to the directional flow of information across the neural matrix, which, in an over-simplified summary, is inaccessible from the input receptor to the central processor[3]. Features of speed, accessibility and so forth can very roughly be grouped alongside the goal of encapsulation with those of operation, pattern and structure coinciding more or less with domain specificity[4]. Where encapsulation is enjoined with the task of delivering unaltered information to its goal, domain specificity provides the dedicated avenues to proceed. The domain specificity of an input system, or module, is the means by which functionally individuated mechanisms can, according to Fodor, generate hypotheses from proximal stimulations. This means that there are dedicated routes for various input systems that gather the incoming data from the external source and map it to the processing center along the appropriate, domain specific routes. The lexicon or convention of that path, in other words, is of the same order as the central processing algorithm; in short, it encodes the information in a way that only it can and

delivers it in a salient format for evaluation. For example, visual perception input systems would map transduced retinal information into a central computational language specific to the retinal coding, whatever that might be. Thusly, the retinal input system has a domain specific task wherein it is expected to fulfill only those sets of informational transfer and encoding obligations in a language recognizable to the central processor.

§3. Poverty of the Stimulus Argument: Nativism and Empiricism/Non-Nativism

Among the arguments proffered by those in support of theories of cognitive modularity[5], Chomsky's poverty of the stimulus (POTS) argument represents a symbolic touchstone for Fodor and Sterelny before parting ways. The POTS argument claims that language learning qua learning is impossible given constraints of time and information available demonstrating instead the innateness of certain language specific capacities within the mind that supplement these deficiencies. One version of the argument runs roughly as follows:

1. Human languages demonstrate infinite recursion patterns of grammar hierarchy.

2. For any group of sentences assembled from this grammar hierarchy (which is capable of infinite recursion) there is an indefinite number of grammar hierarchies that could likewise have been produced to express the same information. Instances of accepted use of grammar (positive evidence) fail to provide enough information such that correct grammar can be learned. Therefore instances where the information in not accepted by the grammar (negative evidence) are needed. Young human language learners are only exposed to positive evidence and not negative evidence (i.e. they are raised hearing sentences of comprehensible language patterns).

3. Young human language learners indeed learn correct grammar usage for comprehensible sentence structures.

4. Therefore, human beings must possess an innate language capacity.

For the nativist, the POTS argument opens the door to the possibility of innate cognitive function and therefore to the idea of modular potentiality. If there exist features of the brain with inherent[6] structures for language learning then inductive reasoning reveals there may likewise exist other components of the brain similarly structured for modular roles. Fodor regards the argument as having little direct impact on a modularity thesis except to posit the conceivability of innateness as a variable of mind[7]. Sterelny, unable to get around the POTS argument with regards to language takes any notion of innateness to be applicable to language alone, declaring that it otherwise plays no role in mental grouping and adds nothing further to modular theses[8]. In short, he takes language learning to be the exception to the rule[9]. It is here that Sterelny outlines a new structure of cognitive organization.

§4. Sterelny: Neural Plasticity, Decoupled Representation and Niche Construction

Forced to circumvent theories of modularity and innateness, Sterelny rebuilds a neural schematic for human cognitive architecture from the ground up by traversing the cognitive evolution of enduring traits including the viable circumstances under which they are obtained. Sterelny's goal is twofold: first to posit niche construction as a contending explanation of human cognitive capacities, and second to determine a position on the eliminativsim debate. This latter goal is a secondary subject beyond the scope of this paper and will be largely ignored. Sterelny's account falls within a functionalist arena where empiricism and evolutionary naturalism intersect. An account of behavioral plasticity and adaptability are what he calls the "wiring-and-connection" facts about human agency. The "wiring" facts are facts about our

internal organization whereas the "connection" facts are "facts about how that organization registers, reflects, or tracks external circumstances" (Stich 2004, p. 492). Sterelny sets about to demonstrate how a cognitive system with an architecture utilizing decoupled representations of goals and targets might have come to be. Decoupled representations are internal states that track facets of an agent's world but are unable to control behavior thus making an agent's actions sensitive to a greater array of information sources (Sterelny 2003, p. 29). Refinement of detection systems during an agent's cognitive development creates a better chance for species survival and thus the organism adapts at the calculated risk of forfeiture of other traits. Multi-cue driven behavior eventually beats out single-cue drives for the advantage it has in increasingly complex circumstances. As paradigms of domain and contest change, transparent models of environmental interpretation are pursued over translucent and opaque options, all lending strength to a broad-banded rather than narrow-banded response breadth in the service of a more robustly developed resource profile. Simply put, neural plasticity offers optimal long-term organism endurance[10] by granting functional neural cooperation via intermodal routing.

Sterelny considers the role of folk psychology, or *theory-theory*, in describing the accuracy of conventional conceptions of mental architecture and posits a dichotomy of two extreme poles between Fodor and the Simple Coordination Thesis on the one side and the Churchlands with their rejection of theory-theory on the other. Fodor's Simple Coordination Thesis posits that our interpretative concepts regarding the organization of human cognition is a putative description of the wiring-and-connection facts that is largely successful and true (Sterelny 2003, p. 6). The Churchlands on the other extreme maintain the opposite, namely that folk psychology is an inutile theory of mind that fails to meaningfully explain mental processes like a robustly mapped future neuroscience will. Theory-theory, according to Sterelny, is committed to the position that categories of belief and preference in some way correspond to organizational features of our

cognitive schema. He goes on to claim that "…we form and use decoupled representations and we form and use representations of the targets of our actions. If nothing in human cognitive systems correspond to beliefs and preferences, then folk psychology does not describe the basic architecture of our cognitive system" (Sterelny 2003, p. 30).

With the framework in place, Sterelny is then able to posit a final operational study before advancing his cognitive model. The Social Intelligence Hypothesis (SIH) serves to bridge the gap between Sterelny's outline of developmental agency and his final theory of niche construction. SIH runs as follows:

1. The increasing demands of social life drive the evolution of agents with intentional psychology: that is, agents with a cognitive architecture characterized by:

    a. Rich representation (these agents track a rich array of factors about their social and physical environment)

    b. Decoupled representation;

    c. In virtue of (a) and (b) the control of action becomes more complex, for the agent notices more about the world, and more of what they notice is relevant to each choice. As a result of the increasing complexity of control, these agents come to have preferences as well as beliefs.

2. As a consequence of (1), the behavior of other agents is less dependent on immediate features of their environment. Prediction of what other agents will do remains critical to fitness but it becomes more difficult.

3. The only plausible solution to (2) is to evolve, despite its costs, a prediction engine: a capacity that enables an agent to predict other agents' behavior reasonably accurately via a reasonably accurate picture of those agents' inner world.

4. Folk psychology is that prediction engine. Our ancestors, perhaps including the last common ancestor of the hominids

and the great apes, probably had it too, albeit in a rudimentary form (Sterelny 2003, pp. 56–57).

From SIH we are able to arrive at the conclusion of Sterelny's presentation. How the cognitive organization of the mind might function given the evolutionary footpath and do so without the aid of modularity is due to what Sterelny identifies as niche construction. Niche construction is a means of altering the environment in future-generation affecting, but non-genetic ways. According to Sterelny, cumulative niche construction modifies the epistemic environment and assembles the scaffolding for the possibility of skill development and cognitive growth, etc. that can otherwise not be acquired. This is a downstream, cumulative epistemic niche construction account that demonstrates the plausibility of natural selection favoring genes that modify cognitive capacities in other ways (Stich 2004, p. 491). Sterelny's point is that despite the numerous internal modifications and adaptations made during the eons of human cognitive evolution in reaction to environmental stimuli and change, SIH and downstream niche construction demonstrate the unique phenomenon of change a socialized mind has on the external world. SIH is an example of niche construction wherein the outside world is changed or manipulated in some way by the cognitive architecture. Such profound abilities aid dramatically in complex evolutionary advances without the need of modularity.

§5. Against Fodor and Sterelny: The Frame Problem

Both Fodor and Sterelny succumb to an epistemological[11] rendition of the frame problem despite independent beliefs of maintaining the nearest semblance to a solution. The frame problem is a long-sustained regress dilemma of knowledge most commonly confronted in AI (and thus to any system of cognitive architecture), the breadth of which deserves considerable attention. A brief recapitulation of the logical form of the frame problem is

as follows: Using logic (mathematical/symbolic), how could one write a formula that adequately describes the effects of an action without being forced to write a yet larger number of resulting formulae that capture/describe the ordinary non-effects of those actions? An epistemological analog of the frame problem can be illustrated through the following short narrative: Imagine a simple robot, one programmed with sentence-representations of the world, lifts a pencil from a table. How upon action completion can the robot refine the scope of resultant propositions about the external world? Following the altered position of the pencil, a complete recalculation of all perceptual variables would be warranted and all internal propositions would thus need to be inspected for potential revision. Succinctly put, the frame problem challenges the ability of a sentient being to cognitively account for a quantitatively undeterminable integer of non-effects such that it can appropriately form corresponding forward-moving mental states. It would simply not know the entirety of the cause-effect sequence, the scope of any and all non-effects, or how it would ever know it had finished considering all of these nebulous variables. This illustration, of course, only concerns a single action by a simple entity. The frame problem proliferates when dealing with organisms of considerably greater cognitive complexity and discerning, not to mention the staggering amount of high speed, constantly transmuting input data. The additional epistemological quandary of how the robot could ever know when and if it had completed the recalculation is particularly troubling because it has to do with knowledge about knowledge. Both Fodor and Sterelny, then, are faced with the difficulty of computational intractability. For Fodor, the issue is one of escaping the problem of framing both within the modular quadrants of cognition as well as in the higher-level processing region. Thus he is faced with the difficulty of each module succumbing to smaller-scale frame problems and the entirety of the module/processor relationship itself being framed. For Sterelny, the problem is less segmented but more challenging as it threatens his overall evolutionarily directed system of cognition.

Fodor's approach, in the heuristic tradition, divides the local from the general and argues that it seems plausible that entire systems of belief may not be as affected as purported: "…There is certainly considerable evidence for heuristic short-cutting in belief fixation, deriving both from studies in the psychology of problem solving and from the sociology of science" (Fodor 1983, p. 115). This division of the modular cognitive components from the central processing umbrella presents a solution for Fodor. The central processor, which Fodor considers to be isotropic (i.e. multi-directional in operative function) and Quineian (i.e. entire-system based), is informationally unencapsulated by virtue of its being non-modular. If the problem is one of complete propositional reformulation then, given the evidence of short cuts in overall belief reactions, the modular organization of our mental architecture aids in circumventing the corruption of the frame problem through a system of localizability. Since the frame problem cannot be solved by domain general systems of cognitive architecture, and the frame problem is clearly solved by us as extant, un-seizing biological organisms of computational autonomy, then modularity must be present and active. An over-simplified formulation of Fodor's argument can be presented as:

(1) The frame problem affects only domain-general cognitive systems that cannot adequately corral the need of unnecessary recalculation.

(2) The mind is modular and, thus, domain specific.

(3) Given this modular cognitive architecture the need for an entire reformulation of belief systems is circumvented and ordinary inertia is possible.

(4) Ordinary inertia occurs, so the frame problem must be solved

(5) (Additional observation): If the frame problem is solved, then the mind is modular.

Sterelny, being a proponent of an even greater degree of domain-general mental organization, has a steeper hill to climb

according to this argument. His maneuver is to attack premise 1. Assuming that the mind possesses, to some extent, a preliminary system of encapsulated information networks and these systems are finite, then some of these systems might contain a large amount of information. "Encapsulated modules within massively modular architectures are threatened with a form of combinatorial explosion. If the mind is massively modular, how are modules coordinated?" (Sterelny 2003, p. 209). Sterelny's question, though posed to a proponent of greater modularity in this instance, is likewise relevant to Fodor. By refining the size of the hard drive that needs to be scoured, how has the problem been sufficiently downsized or removed? The frame problem creates something of a combinatorial explosion, or information collision that is unmitigated by an overwhelmed central analyzer. But Fodor's hybrid schema of modular/central organization is such that given a relatively small number of modules there must be an inverse quantity of encapsulated information within, all of which is susceptible to the same explosion. He continues: "Without an account of coordination and task assignment, there is no reason to suppose that the idea that the mind is a complex of interacting modules avoids the frame problem. Like Fodor, I am persuaded that the frame problem is real, but that somehow or other we solve it. We manage to prevent the combinatorial explosion of possibilities from overwhelming our inferential capacities. Whether human minds are ensembles of modules or not, we are stuck with some version of the frame problem" (Sterelny 2003, pp. 209–210). Sterelny's approach to the frame problem ends here to little satisfaction. By merely pointing out Fodor's shared fate he has failed to succeed himself. One possible direction Sterelny might consider moving in is using his niche construction theory to show that by altering the external world through socialized evolutionary cognition, a different perceptual and input data relationship than normally addressed by the frame problem could somehow be cultivated. Sterelny's combination of theory-theory and niche construction theory could potentially be used to create an argument for something of an epistemological loop or union between mind and world that somehow

skirts the conventional frame problem perimeters. If, according to Sterelny, the socialized mind somehow effects the environment and the non-effects of the environment create the beginning of the problem of mental framing, then Sterelny is poised to develop at very least an argument based on this unique relationship that could move us a little closer to a solution. Instead he simply considers the frame problem to be a non-issue since we function normally without the puzzle impeding daily life.

The frame problem represents seemingly endless difficulty for computational theories of mind and an enormous amount of scholarship has been devoted to making sense of the trap. In comparing the models put forth by the philosophers in question a modest attempt to reframe the frame problem will be put forth here. Sterelny's criticism of a Fodorian solution seems valid. Modules would indeed fall victim to the same threat as the central processor in terms of data quantity needing parsed. Fodor's informational encapsulation only removes this parsing problem one step back to the central processor again. Because the frame problem, true to its name, frames the closed system entirely, either the central processor or the hybrid processor/module combination requires surveying. It seems the available moves to make are to either attack the frame problem as being a syntactically entrapping puzzle[12] or to form a description of the problem as a presently irreconcilable truth[13]. This second avenue is the one seemingly adopted by both Fodor and Sterelny, and for good reason. It seems to me, however, that combining the two would make a stronger argument. We know that the frame problem, when described, seems convincingly immobilizing. Yet we find ourselves operating normally without ever giving it much thought. The problem that needs to be addressed is the variable of non-effect and the epistemological ramifications of trying to input such data. Given the existence of cognitive and perceptual error, it seems clear that whether or not we unwittingly attempt to reanalyze perceptual situations and all the information entailed, we rarely do so infallibly. This at very least suggests a limit in the extension of our thought toward given actions. Likewise, we certainly never attempt to consider

the non-effects of actions performed outside of ourselves (i.e. by other beings, objects, etc.) because the amount of information to parse is unquantifiable. As such, we can at very least posit some kind of cognitive latch that simply disallows the performance of unnecessary automated mental processing. The syntactic trap of the frame problem is the component of non-effects, a variable, it seems, that need not enter the algorithm. The irreconcilable truth is that despite allowing this variable to enter logical philosophical discourse, sentient beings still operate as if unframed. Whether or not this is a stab in the right direction, reconsidering the concept of non-effect seems at very least a promising start.

§6 Empirical Research and the Poverty of the Stimulus Argument Revisited

Fodor relies on the experimental data of sensory systems to bolster his claims of modularity. Certain perceptual responses demonstrate a degree of reactive autonomy apart from any central method of cognitive computation. Phonetic acoustic experiments reveal a domain specificity of auditory deciphering, he claims, noting that since such inputs are informationally encapsulated (and therefore unaffected by pre-existing cognitive information) isolating phonetic speech from non-speech within a sound stream is not only possible but it actually becomes impossible to *not* hear the utterance as an utterance. This, he posits, demonstrates the domain specificity of the utterance/particular language input system module as well as the mandatory operation of the component. Given this inherent modular organization we can understand why visual input system experiments like Müller-Lyer illusion lines[14] continue to mislead observers even after conscious awareness of their equality (Fodor 1983, p. 66). Modularity, for Fodor, outlines a working relationship between mental architecture and mental content. “These results are compatible, so far, with the view that a great deal of the developmental course of the input systems is endogenously determined… linguistic performance—though obviously not present in the neonate—appears to develop

in an orderly way that is highly sensitive to the maturational state of the organism, and surprisingly insensitive to deprivation of environmental information" (Fodor 1983, p. 100). This notion of innate cognitive function returns the discussion to the poverty of the stimulus argument for analysis.

Fodor's nativist position accounts for Chomsky's language acquisition device with a modular and more or less innate function where Sterelny takes a different and less clear path. "We clearly have the potential to 'automate' cognitive skills not subserved by purpose-built wetware… The poverty of the stimulus argument does not sustain some of the central hypotheses of evolutionary psychology" (Sterelny 1999, p. 330). Sterelny contradictorily grants this innate function but denies a modular cognitive model possessing the innateness facet. If innateness is a quality of modularity and Sterelny is working against a thesis of endowed neural structure and toward one of evolved cognition, then to concede mental automation jeopardizes his strategy. Sterelny thinks that the defenders of the Simple Coordination Thesis (i.e. Fodor and defenders of theory-theory) are really relying more on an argument from success than the POTS argument. An argument from success claims that the successful use of our interpretive concepts in ordinary interactions show that those concepts appropriately describe our cognitive architecture. Sterelny's only response to a position like this, however, is that there may exist avenues of predictive success that don't rely on a traditional wiring and connection set of facts (Stitch 2004, p. 494). Rather, predictive efficiency might depend on various cognitive adaptations for interactive interpretation. This, as Stich points out, is not really an argument for behavior prediction without a true theory-theory but merely an example of how the Simple Coordination Thesis won't work. Sterelny later confesses that the Simple Coordination Thesis depends heavily on an argument from success and does not reject the argument. So while claiming that the folk have not got it all correct, he does not clearly show what they have incorrect. His proposition of niche construction as a means of cognitive adaptation influencing the social environment (and the environment

in turn influencing neural plasticity) does not offer the leveled response needed.

Fiona Cowie offers a response to the POTS argument against Fodor, which indicates a need to depart from the internalist (nativist) vs. externalist (empiricist) dichotomy. She redefines the issue by analyzing it as representing conflicting explanatory interests[15]. Cowie seems to think that nativists and empiricists somehow aren't debating a particular contrast but her maneuver to shift the definition of the problem presents few resolutions with regard to innateness being a distinct variable of the mind[16]. Like Sterelny, Cowie attempts to reveal that the POTS argument does not really apply to Fodor and that he is actually relying on something else—this time an Impossibility Argument. An Impossibility Argument claims that the acquisition of concepts depends on a preexisting capacity of representation and that it is impossible to learn about the world without the ability to form hypotheses (which is impossible without a system of representation that allows the formation of hypotheses, etc.). For Fodor, the assimilation of complex linguistic capacities in the mind demonstrates a cognitive propensity toward something reminiscent of a universal grammar theory wherein various linguistic principles are held to be innately demonstrable during the ontogeny of human development. Rather than propose that such specified propositional contents are innately *known* given the architecture and function of lower level modularity, Fodor proposes that they are somehow innately *cognized*. While Fodor and Cowie spar over the definition of innate and the distinctions between POTS arguments and those of impossibility, the issue remains: does the mind have pre-established architectures of modularity that make language learning innate? Fodor's resolution of an innate ability to learn as opposed to an innate, pre-formed knowledge seems really only to be a reformulation of the same problem (or the same problem one step back). In a way, this reduces either the knowledge of language or the process of language learning (or any rendering of the innate, domain specific action of a module) to be something of a brute fact. When described as such it seems categorically wrong to say

that something as complex as the learning of intricate syntactic/grammatical languages is a brute-causal process.

§7 Conclusion

In deciphering the architecture of cognition, at least along the lines of Fodor and Sterelny, it does not seem to be the case that it is an either-or proposition with regard to modular and non-modular models. Neither Sterelny nor Fodor adhere to a strict side: the supposed nativist acknowledges the informational unencapsulated central processor and the non-nativist grants the innateness of (at very least and to some unclear extent) a language learning capacity. Sterelny, in pointing out what theory-theory has right fails to indicate where folk psychology has gone wrong. His niche construction theory is, of course, highly interesting and makes for an exciting alternative to previous evolutionary biological accounts, but the gaps in explanation with regard to language learning and POTS arguments still leave the door open for potential modularity. While Sterelny effectively deconstructs massive modularity we come out with a still less than clear idea of just how the mind is organized on his account. Conversely, Fodor fails to outline exactly how the central processor and higher-level cognition functions operate (in spite of a fairly comprehensive mapping of the input system schematic). Despite this, my inclination is that Fodor is somehow closer to solving the frame problem: breaking the dilemma up into smaller units seems somehow intuitive, but just how that is or how it works is not clear. Broadly, Sterelny's cognitive model is a bit more instinctive given the scope of evolution: should the mind have developed in response to competition and hostility, then the more higher-level interaction with input system information the better: having more resources on the ready would be all the more advantageous. A hybrid model of cognitive design indeed seems vertically structured and requires the speed of modularity and the breadth of non-modularity. A system of unencapsulated input channels, however, definitionally disqualifies modularity. As such, a reformulation of the hybrid

model is perhaps warranted. Dedicated input system encoders seem unavoidable given the different hardware possessed by the various sensory uploaders. More (if not many) medial breaks in the path would allow unencapsulated access by the higher-level cognitive domains for subsequent neural routing. This simple model combines the primary function of domain specificity with a many-checkpointed map for various processing events. It also has the potential to limit the scope of frame problem reanalysis by presenting numerous smaller frames within a larger one (as Fodor attempted to do). A second and much less tenable cognitive model might be drawn from Henrik Walter's vision of higher-level cognitive domain. For him, attitudes and beliefs are to be understood as sophisticated adaptive brain states, which are part of a neuronal net or a multidimensional phase space[17]. This concept might apply to lower level input systems as well, eliminating the Fodorian modules and replacing them with an as-yet unexplainable neurophysics. These two models, when placed side by side, begin to look like rough analogs to Fodorian and Sterelnian architectures respectively. They both succumb to the criticisms above as well. The practice of constructing the foundation of cognitive architecture is clearly one that needs to follow the scientific data yielded by neurobiology before serious philosophical progress can proceed.

Notes

1. Or perhaps beter stated as 'modular through and through.'
2. "A final word about Gall. It seems to me that the notion of a vertical faculty is among the great historical contributions to the development of theoretical psychology" (Fodor 1983, p. 22).
3. Instances of accessibility can arise during medial breaks in the chain between the input system and the central processor.
4. This is a rough and limited grouping for the sake of summary brevity.
5. The poverty of the stimulus argument is often used in defense of Massive Modularity but is also used to support variant theories of cognitive modularity.
6. "Innate" and "inherent" are used interchangeably to mean the same thing.

7. Domain specificity and information encapsulation, according to Fodor, are easily divisible from any notion of innateness allowing for single stimulus mechanisms to be fully encapsulated, general learning mechanisms with the added quality of being somehow inherent. That is to say, Fodor does not lean so heavily on the POTS argument for his modularity thesis, but nevertheless the symbolic role of the POTS argument is one of a dividing line between the nativist and empiricist approach to cognitive architecture.
8. Ironically, the argument intended to stand as the counterpoint between the two philosophers instead reveals a semblance of agreement. Granting this peculiarity, the POTS argument is invoked merely as a reference for the larger division between nativist and empiricist.
9. The *only* exception to the rule.
10. i.e. survival.
11. And, by association, the logical rendition of the frame problem. The frame problem in modal logic is an issue of symbolically capturing the vast non-effects presented by the dilemma. The epistemological version of the frame problem is described above.
12. Perhaps akin to a Zenoian puzzle or a Wittgensteinian argument against linguistic ensnarement.
13. Along the lines of the free will debate, etc.
14. Müller-Lyer illusion lines are parallel lines with opposite facing hash marks or chevrons at either end which, at a glance, seem to show a difference in line length when in fact they are of equal length.
15. For example, accounts for the developmental features of a newborn as debated by a geneticist and an epidemiologist would yield less substantive accounts than would those of a molecular biologist and epidemiologist, the latter combination having more causal relevance (Cowie 1999, p. 22).
16. "I've been arguing that the position defended by Fodor… is of a kind with the radical nativisms propounded by Leibniz and Descartes three centuries ago. Like theirs, Fodor's view—whether disguised as an innateness hypothesis or not—is indicative of a certain degree of pessimism as to our prospects for understanding how a mind comes to be furnished" (Cowie 1999, p. 111).
17. The prefrontal cortex generates future scenarios and the amygdala and hypothalamus affect the body viscerally. The results of these agent-unique simulations play an active role in the adaptive decision-making process of the brain (Walter, 2002).

Works Cited

Carruthers, Peter. (2006) *The Architecture of the Mind: Massive Modularity and the Flexibility of Thought* (Oxford: Clarendon)

Cowie, Fiona. (1999) *What's Within?: Nativism Reconsidered* (New York: Oxford University Press)

Fodor, Jerry A. (1983) *The Modularity of Mind: an Essay on Faculty Psychology* (Cambridge, Mass.: MIT)

Robbins, Philip "Modularity of Mind", *The Stanford Encyclopedia of Philosophy* (Summer 2009 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2009/entries/modularity-mind/>.

Shanahan, Murray, "The Frame Problem", *The Stanford Encyclopedia of Philosophy* (Winter 2009 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2009/entries/frame-problem/>.

Sterelny, Kim, and Paul E. Griffiths. (1999) *Sex and Death: an Introduction to Philosophy of Biology* (Chicago, Ill.: University of Chicago)

Sterelny, Kim. (2003) *Thought in a Hostile World: the Evolution of Human Cognition* (Malden (Mass.): Blackwell Publ.)

Stich, Stephen. (2004) "Some Questions From a Not So Hostile World," *Australasian Journal of Philosophy* 82.3, 491–98

Walter, Heinrik. (2002) "Neurophilosophy of Free Will," in: Kane, Robert, *The Oxford Handbook of Free Will,* pp. 565–576 (Oxford: Oxford University Press)

For a Lovelier Understanding in Explanation[1]

James Winburn

In this paper, I take a critical stance against the philosopher of science Nancy Cartwright's proposal for a future program in scientific explanation that advocates a particular causal model. My goal is to reveal why Cartwright's universalist model on causation fails to provide an adequate explanatory program for science. My criticism takes aim at her 2003 article, "From Causation to Explanation and Back," while encouraging a critical reconsideration of Cartwright's proposal for a universalist model of causal diversity.

I support the same explanatory power that Cartwright aims for in her own model, a criterion that I will define in terms of a worthy goal of inquiry capable of addressing the different areas in science. But I will argue that causation, or a causal model alone, cannot serve as a tenable program of explanation that is capable of addressing these needs. My claim is that if our aim in scientific explanation is for the sort of diversity that Cartwright intends, one that covers the many different areas of science, then it cannot be achieved with a singular model. I will argue that only a pluralistic approach to explanation can address this need. A universalist approach to causation will invariably neglect giving proper attention to theoretic issues that fall outside causal schemes, such as in the case of areas within the social sciences. I will conclude by showing the growing need for support toward pluralism as a future program in scientific explanation.

My argument for a pluralistic attitude in scientific explanation addresses a set of explanatory criteria based on a worthy goal of inquiry. As a general note, the word 'pluralism' refers to the view that there are many things in question, such as concepts, scientific world views, discourses, viewpoints, etc. (Mason 2006).

I will use the term 'pluralism' to denote whatever theoretical model (whose form is not a singular and universal prescription in scope) intended to adequately address the various explanatory needs of different areas in science. Such a model of pluralism will be characterized later in this paper by Peter Lipton's challenge for an explanatory construct that addresses the greatest degree of understanding—the *most "lovely"*—rather than the most probable explanation available.

Philosopher of science Peter Godfrey-Smith advocates a pluralistic model in scientific explanation where there is no single theory of explanatory relation. He foresees the emergence of a definite kind of pluralism about explanation in the philosophy of science. According to Godfrey-Smith, the most peculiar thing about the discussion of explanation by philosophers has been the assumption that explanation is the kind of thing that requires analysis in terms of a single special relation or a short list of special relations. He says that it is a mistake to think there is one basic relation that is *the* explanatory relation (as in the covering law theory, the causal theory, and the unification theory), and it is also a mistake to think that there are some definite two or three such relations (Godfrey-Smith 2003).

The alternative view that Godfrey-Smith recognizes is that the idea of explanation operates differently within different parts of science, as well as differently within the same part of science at different times. He maintains that what exactly is being sought in explanation is not constant in all of science. "If an *ism* is required, the right analysis or explanation is a kind of *contextualism*, a view that treats the standards for good explanation as partially dependent on the scientific context," he says, and notes that Thomas Kuhn argued in his 1977 article, "Concepts of Cause in the Development of Physics", for a view of this kind (Godfrey-Smith 2003). In his paper about the history of physics, Kuhn claimed that different theories (or paradigms) tend to bring with them their own standards for what counts as a good explanation.

Pluralism is inherently different from universalism in that a universalist model, such as Cartwright's, asserts that there is

some singular form shared across the sciences that a good theory should take up (Cartwright 2003). However, Cartwright seems to be appealing to a pluralistic ideology for the construction of her own universalist program. She notes that the science community has seen a number of accounts of causal-explanatory relevance, while realizing that there are also a variety of different accounts of singular causal explanation on offer. When faced with the question to which account may be the correct one, Cartwright says that they probably all are, prescribing that each one serves the need of a different kind of causal relation (Cartwright 2003). Cartwright may be suggesting a form of pluralism here, but only in that it pertains to a universalist model based exclusively on the varieties of causation. This is the fundamental difference between the open-program variety of pluralism I am advocating. Such a pluralistic program takes many different models into consideration (not only causal types), thus going beyond causal inference to address explanation in the various fields of sciences. As I will show, this seems the more likely solution to our definition of a worthy goal of inquiry than for Cartwright's universalist construction that is limited to the strictures of causal models.

The conclusion that Cartwright draws for a future program in scientific explanation is the need for causal diversity. She says she is a strong advocate of causal diversity at the level of general causal relevance, which is in reference to a variety of different kinds of "singular causal relations" that can obtain, such as hasteners, delayers, sustainers and contributors, for example (Cartwright 2003). These refer to different kinds of causal systems and different ways in which a cause can operate within them. Briefly stated, Cartwright's model of causal diversity is formed on the basis that these different methods for testing are appropriate for different specific kinds of causal hypotheses relative to different sets of background assumptions about the system in which the putative causal relation is imbedded (Cartwright 2003). To proceed in this way, Cartwright explains that we need far better accounts of the kinds of causal systems we may encounter and the variety of ways that a cause may operate within them. She

believes her model to entail a vast new project in philosophy of science once we recognize that there are a great variety of causal relations and a great variety of causal systems, of which each may have its own way of testing.

Cartwright explains that her view stands in contrast to how we are used to thinking of causation, that is, the normal assumption that causal relations all have but a single feature in common that distinguishes them from mere association. She cites support for her view from Christopher Hitchcock, who advises that "the goal of a philosophical account of causation should not be to capture *the* causal relation, but rather to capture the many ways in which the events of the world can be bound together" (Cartwright 2003).

The explanatory power that I will outline next is similar to what Cartwright advocates for her own program, addressing a criteria for explanation that may be applied to the various areas of science. This notion for an adequate approach to explanation must first recognize what criteria a worthy goal of inquiry should hold for our needs of understanding in explanation. What I believe to be central to this type of explanatory criteria is a pluralistic program of explanation that adapts itself to the many different areas of science, whether theoretically dependent upon causal processes or paradigms that defy causal relationships.

James Woodward captures this necessary sense of explanation by focusing on how explanation itself relates to explanatory goodness, evidential support and other goals for inquiry. Woodward says that most writers on explanation have not always given adequate attention, beyond our intuitive judgments, to how explanation itself is connected to, or is even distinct from, other goals of inquiry—such as the connection between explanatory goodness and other frequently proposed goals for inquiry in terms of evidential support, prediction, control of nature, simplicity, and other virtues (Woodward 2003).

To achieve an adequate criterion for explanation, Woodward says we should focus on the value of the information that an explanation provides and how exactly it relates to the goals we value in our inquiry. It is this attention toward informational output that

Woodward says should be used to identify and address the different explanatory needs of the various areas of science. According to Woodward, "One way forward in assessing competing models of explanation is to focus less (or not just) on whether they capture our intuitive judgments and more on the issue of whether and why the kinds of information they require is valuable (and attainable)" (Woodward 2003). Woodward's reference to intuitive judgments here refers to the general tendency toward assigning causal paradigms to our need to explain scientific phenomena.

Woodward considers Wesley Salmon's Causal Mechanical (CM) model as an illustration. Underlying the CM model is presumably some judgment to the effect that tracing causal processes and their interactions is a worthy goal of inquiry. One might try to defend this judgment simply by claiming that the identification of causes is an important goal and that causal process theories yield the correct account of cause. But a more illuminating and less question-begging way of proceeding would be to ask how this goal relates to other epistemic values found throughout the different areas of science. Woodward challenges us to consider the connection between the goal of identifying causal processes and constructing unified theories, or between identifying causal processes and the discovery of information that is relevant to prediction or to manipulation and control when testing scientific theories (Woodward 2003).

According to Woodward, when we construct an explanatory model based on the value of information it provides us, we are faced with the question of whether or not these are the same goals, or whether they are independent but complementary goals. If so, we need to consider whether they are competing goals in the sense that satisfaction of one may make it harder to satisfy the other. Woodward says that one may ask similar questions about the goal of other explanatory models, which are universalist in construction, such as Carl Hempel's covering law theory, Salmon's causal theory, or Philip Kitcher's unification theory[2] (Woodward 2003).

Peter Lipton's Inference to the Best Explanation thesis supports Woodward's argument for a worthy goal of inquiry,

implying that we are inevitably faced with many possible models of explanation to justify a suitable criterion that spans the different areas of science. Lipton points out key characteristics of a worthy goal of inquiry to help bring to light what will satisfy such a comprehensive program of understanding in scientific explanation. According to Lipton, emphasis must be given to a worthy goal of inquiry in constructing a model of scientific explanation if it is to maintain informative value (Lipton 1991). Without emphasis toward a worthy goal of inquiry, one faced with various choices of explanation will inevitably construct a model of explanation that is most "likely", as opposed to most "lovely", which, according to Lipton, severely reduces the interest of the model by pushing it towards triviality. What Lipton is referring to here is the process of constructing a model that fits our expectations of what we believe to be informative (taken as "most likely"), rather than constructing a model based on a worthy goal of inquiry (taken as "most lovely"). (Lipton 1991)

Lipton says that models of explanation normally yield disappointing results when they yield to this process of Inference to the Best Explanation. When considering a basic question about the sense of "best" that the model requires, Lipton challenges us to consider whether we mean the most probable explanation ("most likely"), or rather the explanation that would, if correct, provide the greatest degree of understanding ("most lovely"). In a more fundamental sense, Lipton says that we need to examine those factors that make one explanation better than another. He says that unless we can say more about explanation, the model will remain relatively uninformative. In the case of the most informative model, Lipton says that the model should be construed as "Inference to the Loveliest Explanation." Lipton means his central claim to be that scientists take "loveliness" as a guide to likeliness, so that the explanation that would, if correct, provide the most understanding, is the explanation that is judged likeliest to be correct (Lipton 1991).

Lipton's thesis raises three challenges for this goal in explanation. The first is to identify the explanatory virtues, that is, the

features of an explanation that contribute to the degree of understanding they provide. The second is to show that these aspects of "loveliness" match judgments of "likeliness," meaning that the loveliest explanations tend also to be those that are judged likeliest to be correct. His final challenge is to show that by granting the match between loveliness and judgments of likeliness, the former is in fact the scientists' guide to the latter (Lipton 1991).

Both Woodward and Lipton provide us with a positive structure of criteria for forming an adequate definition of explanation that addresses scientific understanding across the sciences. With this criteria in mind, Cartwright's advocacy for causation as a basis for her program faces the traditional problems inherently found in causal inference. Her 2003 article tracks the shift from causation to explanation, outlining the troublesome relationship between the two, while ultimately tracing the gradual transformation of explanation back to the tendency toward causation.

Cartwright says that Hempel's Deductive-Nomological (D-N) model was recognized from the start as too restrictive. Insistence on deduction from strict laws would rule out a large number of explanations in which we place a high degree of confidence. Hempel then added the Inductive-Statistical (I-S) model, where explanations may cite law claims and individual features from which it follows that the facts to be explained are highly probable even if they do not follow deductively. The underlying idea in both cases (D-N and I-S models) is that the factors offered in explanation (the explanans) should provide good reason to expect that facts to be explained should occur (the explanandum).[3] Cartwright says this can be traced back to David Hume's account (*An Inquiry Concerning Human Understanding,* 1748) in which the idea of causation is a copy, not of an impression of a relation in the external world, but rather a copy of an impression of our own feeling of expectation that the effect will occur when we observe the cause. For Hume this feeling of expectation results from our repeated experience of the cause being followed by the effect. Contending with this explanatory obstacle, advocates of the D-N and I-S models believed that expectation needed to be rational,

not habitual, that is, it must be justified under well-confirmed law claims (Cartwright 2003).

According to Cartwright, Wesley Salmon's Statistical Relevance (S-R) model was an attempt to capture the features of causal relevance that appeared to be left out of the empiricist constraints in an explanatory model. Salmon set out to capture these features in terms of statistical relevance or conditional dependence relationships. His proposal was that the explanans should be statistically relevant to the explanandum, meaning that the probability of the explanandum should be different when the explanans obtains from when it does not. However, Cartwright notes that the various universalist models failed to overcome the Humean challenge to circumvent our habitual expectations through rational justification, whether through well-confirmed law claims or statistical relevance (Cartwright 2003).

In the eager attempt to rid science of the traditional concept of causality, Cartwright notes that one standard strategy has been to "reduce it away," that is, to eliminate all use of causal notions and define causation purely in terms of "regular association" plus some other non-causal concepts like "temporal succession" and "spatio-temporal contiguity." One such attempt was Salmon's follow-up to the SR model, his Causal Mechanical (CM) model that set out to address causal and explanatory relationships over and above facts about statistical relevance, while remaining within a Humean framework (Cartwright 2003).

Salmon himself states the problems that he encountered with the search for explanatory relationships are causal in nature, saying that the problem is the universalist tendency to make a singular causal program applicable across the span of science's different areas of research. He says, when we think seriously about the very concept of scientific understanding, it does not seem plausible to expect a successful characterization of scientific explanation in terms of any simple formal schema or simple linguistic formulation (Salmon 1990). He notes that an answer may be found in the two virtues that require more attention: the first, in terms of unification; and the second, in terms of exposing underlying mecha-

nisms that will satisfy causal explanation. Salmon notes that it may be a futile endeavor to try to explicate the concept of scientific explanation in such a comprehensive manner. He recommends that it may be a more worthy enterprise to list various explanatory virtues that scientific theories are said to possess and to evaluate scientific theories in terms of them. Regardless, Salmon is certain that future research in scientific explanation must take as its motivation the search for additional scientific explanatory qualities (Salmon 1990).

Cartwright's program as a universalist model becomes problematic due to this inability for a singular model to address the different needs of the different areas of science. Her universalist approach to causation may intend to address the above explanatory criteria through causal diversity, but it is inevitably confined to the strictures of a rigid model of explanation. She says that recognizing this should make us more cautious about investing in the quest for universal methods for causal inference (Cartwright 1999). She acknowledges that it is more reasonable to expect different kinds of causes operating in separate ways or imbedded in differently structured environments to be tested by separate kinds of statistical tests. Her particular model of causal diversity takes this criteria as its explanatory hallmark for better explanations that explain more types of phenomena, to explain them with greater precision, and to provide more information about underlying mechanisms.

However, Lipton notes that some of these features that pertain to a particular model have proven surprisingly difficult to analyze because the model of causation cannot by itself go it alone to support these many characteristics of explanatory power. A universalist model, such as Cartwright's, fails in its explanatory power only because it neglects to consider other non-causal explanatory models that may address theoretical issues that a causal model cannot address. In addition to "temporal succession" and "spatio-temporal contiguity," another such example of a non-causal explanation is an idea advanced by Graham Nerlich ("What Can Geometry Explain?" 1979), known as "geometrical," which

is non-causal in the sense that it explains phenomena by appealing to the structure of space-time rather than to facts regarding the transfer of forces or energy/momentum between objects or events (Woodward 2003).[4]

Noting that explanation can hardly be limited to causal processes alone, Woodward says that to completely subsume explanation into causation loses connections with important issues involving our explanatory criteria (Woodward 2003). Cartwright's prescription for handling a future program in scientific explanation falls short of its explanatory strength because she advocates a one-size-fits-all solution to be used across the board in science. And considering the traditional problems with causal inference, a universalist model of a causal variety cannot adequately serve this purpose in scientific explanation. Woodward argues that causal claims themselves seem to vary greatly in the extent to which they are explanatorily deep or illuminating. He explains that even if one focuses only on causal explanation, there remains the important project of trying to understand better what sorts of distinctions among causal claims matter for "goodness in explanation." To some extent, this is so; the kinds of concerns that have motivated traditional treatments of explanation do not seem to be entirely subsumable into standard accounts of causation (Woodward 2003).

If causation is to play any role in explanation beyond our intuitive judgments, then it must do so in a pluralistic sense requiring the aid of other types of models. In such a way, other models may address causation's shortcomings to complement the explanatory needs of science in its various areas. Woodward addresses this need for pluralism in explanatory models, saying that the only effort in the future will be devoted to developing models sensitive to disciplinary differences. Noting that past programs in scientific explanation have generally aspired to be universalist, Woodward questions the extent to which it is possible to construct a single model of explanation that fits all areas of science. He maintains that it is uncontroversial that explanatory practice varies in significant ways across different

disciplines, taking into consideration what is understood by an explanation, how explanatory goals interact with others, and what sort of explanatory information is thought to be achievable, discoverable, and testable (Woodward 2003).

Woodward maintains this position in light of the extreme view that explanation in biology or history has nothing interesting in common with explanation in physics. He says that ideally such models would reveal commonalities across disciplines but they should also enable us to see why explanatory practice varies as it does across different disciplines and the significance of such variation. For example, Woodward says that biologists, in contrast to physicists, often describe their explanatory goals as the discovery of mechanisms rather than the discovery of laws. Although it is conceivable that this difference is purely terminological, it is also worth exploring the possibility that there is a distinctive story to be told about what a mechanism is, as this notion is understood by biologists, and how information about mechanisms contributes to explanation (Woodward 2003).

In this paper, I have stressed the need to widen our scope of explanation beyond the confining framework of causation as a singular and exclusive model for scientific explanation. It is my claim that an explanatory program should take as its goal a model of pluralism if it is to adequately serve the need to explain across the sciences. Such a pluralistic model is most effectively delineated by what Lipton refers to as a model whose criteria is based upon a worthy goal of inquiry (the *most lovely*), rather than one that simply fits our expectations of what we believe to be informative (the *most likely*). Though Cartwright takes into consideration this explanatory virtue for her model of causal diversity, she confines her explanatory power to a narrow construct that fails to open itself to the advantage of considering models outside the scope of causation, such as "geometrical" or other non-causal forms of explanation.

Even though I have asserted pluralism as the most comprehensive approach to explanation, I believe we should always put forward the reason why we are clear about the assumptions under

which a particular model is valid and those under which it is not. Even if we should hold one explanatory model to be superior to another, we should not use it as an argument to relieve ourselves from describing explicitly how our assumptions may hold in a particular theoretical case. Instead, consideration must always be given in the different disciplines of science on just how a particular model may prove to be instrumental toward a worthy goal of inquiry.

Notes

1. Thanks to Dr. Ricardo Gomez, Dr. Steven R. Levy and Dr. Ronald Houts for their guiding input in helping to shape the direction of this paper.
2. The theory on unification, advocated by Philip Kitcher, is the claim that theories unify our knowledge by providing us with certain patterns of argument that can be used to construct explanations. Kitcher's basic strategy is to show that the derivations we regard as good or acceptable explanations are instances of patterns, that taken together, score better than the patterns instantiated by the derivations we regard as defective explanations. Woodward explains that a fundamental problem with Kitcher's unificationist model is that the content of our causal judgments does not seem to be derived from our efforts at unification. For example, considerations having to do with unification do not by themselves explain why it is appropriate to explain effects in terms of their causes rather than vice-versa (Woodward 2003).
3. From Hempel's Deductive-Nomological Model, a scientific explanation consists of an explanandum, a sentence that describes the phenomenon to be explained, and the explanans, which are those sentences that are adduced to account for the phenomenon. If the explanans are to successfully explain the explanandum then, according to Hempel, the explanandum must be a logical consequence of the explanans, and the sentences that constitute the explanans must be true (Woodward 2003).
4. Graham Nerlich asserts that there is a variety of physical explanation, which is "geometrical" rather than causal. Explaining the trajectory followed by a free particle by noting that it is following a geodesic in spacetime is an example of a geometrical rather than a causal explanation, according to Nerlich. Woodward notes that because there are non-causal forms of explanation, scientific explanation will continue to be a topic that is somewhat independent of causation (Woodward 2003).

Bibliography

Cartwright, Nancy. (1999) "Causal Diversity and the Markov Condition," in: B. Skyrms (Ed), *Statistics and Causation*, a special edition of *Synthése*, pp. 3–27

Cartwright, Nancy. (2003) "From Causation to Explanation and Back," in: J. Reis (Ed), *Causality: Metaphysics and Methods*, Technical Report CTR 09–03, Centre for Philosophy of Natural and Social Science, LSE

Godfrey-Smith, Peter. (2003) *Theory and reality: an introduction to the philosophy of science* (Chicago: University of Chicago Press)

Lipton, Peter. (1991) *Inference to the Best Explanation* (New York: Routledge)

Mason, Elinor. (2006) "Value Pluralism." *Stanford Encyclopedia of Philosophy*. (20 June 2006). 28 March 2010. <http://plato.stanford.edu/entries/value-pluralism/>

Salmon, Wesley. (1989) *Four Decades of Scientific Explanation* (Minneapolis, MN: University of Minnesota Press)

Salmon, Wesley. (1990) "Scientific Explanation: Causation, and Unification," in *Crítica: Revista Hispanoamericana de Filosofía*, Vol. 22, No. 66, pp. 3–23 (Universidad Nacional Autónoma de México)

Woodward, James. (2003) "Scientific Explanation." *Stanford Encyclopedia of Philosophy*. (16 January 2009). 17 December 2009. <http://plato.stanford.edu/entries/scientific-explanation/>

A Formal Critique of Intersectionality: Bridging the Gap Between Social Science and Feminist Theory

Jasmine Hall

I. Abstract

Modern feminist philosophy has introduced a doctrine called 'intersectionality,' which proposes to change the way in which we view the sociological aspect of female experience. This view acknowledges that, "subjectivity is constituted by mutually reinforcing vectors of race, gender, class and sexuality" (Nash 2008; p. 2). Although the utilization of intersectionality as an analytical framework has changed the stride of feminist philosophical, sociological and legal thought significantly and necessarily, it has defects that require attention.

In this paper, I will offer a resolution to the flaws that render intersectionality epistemically insufficient as an analytical framework. Useful for viewing the female experience both socially and philosophically, intersectionality has promise as an analytical tool; however, significant modification would improve it by making it more efficacious as a means of promoting greater social equality (Nash 2008; p. 3).

I will begin laying the groundwork by explicating the need for intersectionality. To do this, it is necessary to introduce the concept of feminist thought, development and practices, as understood in the field of philosophy in order to stress the impact of the institutional teachings of feminism on our society. The selective teaching of some concepts to the exclusion of others has led to a need for modification regarding the way in which the female

experience is viewed: enter intersectionality. [1]

Second, I will offer a description of what intersectionality is, particularly as described by legal scholar Kimberle Crenshaw (Nash 2008; p. 3). [2] Crenshaw's formulation of intersectionality is the first and most prevalent and, although philosophers and sociologists have suggested modifications to her explanation of the female experience, her formulation still has a basic foundational nature about it that has endured in spite of disputes among philosophers and sociologists (Collins 1998, p. 63).

Third, I will point out the advantages of intersectionality. Intersectionality accepts the fact that female experience is not equal along cultural, racial and social lines and, thus, attempts to redirect the way in which we view the female experience (Crenshaw 1989; p. 140). It correctly incorporates social and empirical truths into the introspective conception of gender, which is an indispensable component of feminist philosophy.

After discussing the advantages of intersectionality, I will address its shortcomings. The thesis of intersectionality is weak because it fails epistemically. It also prioritizes the black female's experience as being the paradigmatic subject for intersectional analysis (Nash 2008; p. 7). [3] Using 'black *women*' paradigmatically, Crenshaw unintentionally sidesteps the fact that the quality of the experience of the black *woman* is not similar. Most importantly, intersectionality emphasizes the gap between sociology and philosophy when the intention is to mend it.[4]

I will conclude by offering some solutions for intersectionality as it pertains to feminist philosophy. My intention is not to offer a substitute framework, but a jumping off point for furthering the theory of intersectionality as an approach within feminist theory.

II. Introduction to Feminist Thought, Development and Practice

It is true that there is not simply one school of feminist thought. However, I am specifically addressing the philosophy of female

existence in western society and, foundationally, the epistemology of the female experience. Historically, there has been a disparity regarding the popularity of feminist concepts as they are taught institutionally and that disparity, I argue, can be addressed through intersectional analysis.[5] In actuality, the experience of women of color is both similar to and different from the experience of white feminists (Crenshaw 2008; p. 149).[6] Traditionally, early feminism and the way it has evolved into contemporary feminism has contributed to a few female philosophical concepts, but I will focus for the most part on two key feminists whose theoretical views dominate the way in which feminism has progressed: Mary Wollstonecraft and Simone De Beauvoir.

Considered the 'first' feminist, Mary Wollstonecraft receives a large amount of academic regard. She is widely taught and her philosophical accomplishments are very well known in the Western academic world. Wollstonecraft discusses female enterprises and experiences very openly and replies directly to the works of her mentor Jean-Jacques Rousseau and the general attitude of the 'republican mother' that overtook Europe with the overwhelming popularity of his work "Emile"[7], during the expansion of the French Empire under Napoleon (Taylor 1992; p. 1).[8] She drives home the general theme of her writings in a response to the violence towards and degradation of the 'woman,' in "A Vindication of the Rights of Woman" in which she states:

> A wild wish has just flown from my heart to my head, and I will not stifle it, though it may excite a horse-laugh. I do earnestly wish to see the distinction of sex confounded in society, unless where love animates the behavior (Wollstonecraft 1792; p. 3).

Historically, following Wollstonecraft as a frontrunner in feminist philosophical thought is Simone de Beauvoir. In "The Second Sex," De Beauvoir provides not only a biological account for our vivacious gendered attitudes and an explanation for why we have modified them significantly from a social standpoint, but she also offers historical, mythical and actual accounts of

the female experience that has shaped the second-class nature of that experience. De Beauvoir focused on the social institutions and standards women are confined to. About the institution of marriage, she writes:

> Marriage has always been a very different thing for man and for woman. The two sexes are necessary to each other, but this necessity has never brought about a condition of reciprocity between them… a man is socially an independent and complete individual, he is regarded first of all as a producer… the reproductive and domestic role to which woman is confined has not guaranteed her an equal dignity (De Beauvoir 1952, p. 425).

De Beauvoir's philosophy regarding the female experience is incredibly fundamental and definitely stems from Mary Wollstonecraft's theory. Feminism as a discipline is characterized via its 'frontrunners', who examined the way in which the female identity constantly conflicts with that of the male identity in any given society and therefore, in any given social institution that society provides. Feminism as a developmental theory began to receive an escalating amount of attention from the time of Wollstonecraft to de Beauvoir to the present.

These women are the historical frontrunners of modern feminist philosophy, and yet neither Wollstonecraft nor de Beauvoir discuss race as an essential component of feminist thought. This implies that gender is independent of other societal factors, that women have similar experiences with oppression and their relationship with males within a society, and that their collective attitudes about their self-image are synonymous with their outward experience.[9] In light of this observation, intersectional theorists and feminists ask a very poignant question: *Why is it that there are no women of color historically regarded as feminist frontrunners*?[10] There are some reasons why this is the case, and they have come to be of increasing importance to feminism as a whole.

The first of these reasons is the existence of a historical racial hierarchy. Wollstonecraft and De Beauvoir (in addition to

many other early feminists) are of European descent, and therefore hail from a Western society which has particularly exemplary attitudes about different ethnic groups, and which is also responsible for developing the theory of 'race' as separate and distinct sub-species of humans (Blumenbach 1776; p. 3).[11]

A specific example of the concept of 'race' first introduced and used by Johann Blumenbach to illustrate the supposed inherent dominance of one group over another was the development of the European attitude with regard to African slavery. This concept of 'race' that Blumenbach asserted to be inherently scientific[12] was largely already existent and based on many negative attitudes regarding Blacks in Europe, beginning most likely with the emergence of the Moors in Spain and developing throughout Europe over time[13]. Slavery was popular in Europe before the sixteenth century, but was a distant historical memory to England (Northern Europe) at that time, although slavery had existed there in the past. The oppressive attitudes based on race that eventually made their way to the Americas from England to justify slavery were not intrinsic until these attitudes took hold further[14] to make the exploitation of Africans justifiable based exclusively *on* race. These attitudes represented the crucial difference that characterized the oppressor-oppressed[15] relationship that persisted long after the abolishment of slavery in America.[16]

From the concept of a 'historical racial hierarchy' follows that of 'racial superiority' being a contributing factor in the prevalence of white feminists in the academic world. The oppressed must necessarily have oppressors and, in Western society, the oppressed are people of color and the oppressors are people of European descent (Crenshaw 1989; p. 140).[17] With this societal concept firmly in place, it is no mere accident that European philosophers (in general, regardless of sex) dominate the world of philosophy. Moreover, and more specifically, white women dominate feminist philosophy.

Due to this racial hierarchy and the prevalence of Western ideology throughout the world, the oppressor-oppressed relationship still exists. The social axis of 'racial hierarchy' embedded in

Western society affects its social institutions and the way in which capitalism and the legal system actually function in the Western world (CRC 1977, p. 1). The advantages of money and resources enable white philosophers to afford the best opportunities; they are more likely to be published and revised, making their theories appropriate for the audience most likely to advance them. This results in their philosophical endeavors becoming immortal. Consequently, those who are educationally disadvantaged experience far fewer opportunities simply because they cannot afford a higher education.

Educational advantage is important because Western society tends to regard education as the chief sign of a person's intellectual capacity. Valid and important philosophical theories, important to the discipline, that are combined with poor writing skills and a lack of any substantial knowledge of classical philosophy due to educational disadvantages, are judged irrelevant. [18] Sojourner Truth, in her 1851 speech "Ain't I a Woman?" highlights intellect as the difference between women being taken seriously by men and their being dismissed. Specifically, men dismissed Sojourner Truth because of her lack of 'intellect' due to her previous slave-class standing:

> Then they talk about this thing in the head; what's this they call it? [a member of the audience whispers 'intellect'] That's it, honey. What's that got to do with women's rights or Negroes' rights? If my cup won't hold but a pint and yours holds a quart, wouldn't you be mean not to let me have my little half measure full? (Truth 1851).

The chief advantage of education is exposure. Exposed not only to different philosophical concepts, educated and rigorously trained philosophers are also exposed to the majority opinions that influence what are to be the most 'accepted' and plausible theories and methods. From a historical standpoint, feminist philosophers like Sojourner Truth suffered from just such educational disadvantage.

In "Mapping the Margins", Kimberle Crenshaw regards the unjust practice of subordinating women of color based on the

aforementioned racial, social, economic and educational historical inequalities that have anonymously existed for them. She offers an explanation that what commonly happens is that the 'white' female experience emerged at some point as the norm and that the 'colored' female experience was completely marginalized. However, the 'colored' female experience is also crucial to an understanding of the female experience because they are women too, and their experience as women of color is different from the experience of the white woman. What is affected is the perception of women of color and this perception harbors the unjust treatment of women of color and the expectation from them to adapt completely to European standards. In Crenshaw's view, we ought to formulate a philosophy regarding the female experience that incorporates the racial aspects of that experience and to acknowledge that there is no 'norm' for the female experience predicated on sex identity alone.

My next project is to illustrate in what way 'intersectionality' is a response to these racial, social, economic and educational injustices that have created such a difficult situation for women of color. Feminism, as characterized by Wollstonecraft, De Beauvoir and others has failed to recognize women of color, perhaps as a result of society being colored by hierarchal relations.

III. Description of Intersectionality as advanced by Kimberle Crenshaw

According to Kimberle Crenshaw, intersectionality is an analytical framework with which to view the female experience metaphorically as an intersection in which racial and gendered aspects of female existence can be seen as the meeting of 'streets', 'avenues', etc. in order to deductively conclude an individual's position in the intersection of those axes of oppression (Crenshaw 1989; p. 149).[19]

In "Demarginalizing the Intersection of Race and Sex", Crenshaw develops her theory of intersectionality by using the 'black woman's experience' as paradigmatic, or as "the starting point"

(Crenshaw, p. 140). Race and gender are not mutually exclusive categories; a black woman's experience is multi-dimensional and the racial and gendered aspects of her identity inform one another and, perhaps, more foundational facets of her existence, namely, her gender and race. Crenshaw focuses on the experience of black women because she believes they historically have been the subject of the most accumulative oppression over time, so their experience is magnified in relation to other women of color (Crenshaw 1989; p. 143).[20] Crenshaw introduces the notions of 'double discrimination'[21] and speaks in terms of 'privilege and non-privilege'[22] to comparatively consider maleness with femaleness and whiteness with blackness that expresses how "black women were [considered] something less than real women,"[23] and introduces violence against women and injustice in the legal system, thus emphasizing the issues out of which intersectionality is borne. (Crenshaw 1989; pp. 149–155).

In "Mapping the Margins", Kimberle Crenshaw develops her theory of intersectionality further by primarily addressing male violence against women and identity-based politics. She uses these two examples to conclude that intersectionality shows that it is incorrect to view the female experience as one that is independent of other factors (Crenshaw 1991, p. 1248). Crenshaw then continues to advance her theory by looking at intersectionality as structural and political. Structural intersectionality reveals the ways in which:

> [T]he location of women of color at the intersection of race and gender makes our actual experience of domestic violence, rape and remedial reform qualitatively different than that of white women. (Crenshaw 1991; p. 1245). [24]

Political intersectionality reveals the way in which "both feminist and antiracist politics have, paradoxically, often helped to marginalize the issue of violence against women of color. Crenshaw's main objective is to introduce intersectionality in such a way that we can see its relevance in *other* social aspects in addition to using it to examine legal injustices against women of color. Ultimately,

intersectionality is

> implicit in certain strands of feminist and racial liberation movements… [it] is the view that the social power in delineating difference need not be the power of domination; it can instead be the source of social empowerment and reconstruction (Crenshaw 1991; p. 1242).

It is Crenshaw's mission to identify the mistakes of past antiracist and feminist frameworks and ideologies, namely that they have "failed to consider intersectional identities [and experiences] such as women of color", which are "frequently the product of intersecting patterns of racism and sexism" (Crenshaw 1991, pp. 1243–1244).

Since feminism has seen virtually no 'colored' faces and antiracism has seen virtually no women's faces, it is the case that commonly both groups marginalize women of color (Crenshaw 1989; p. 140). Men are historically the leaders of the antiracist movements that receive recognition and advancement. Similarly with feminism, it is white women of Western European descent[25] that are the most recognized in the feminist movement. Patriarchy and its prevalence in Western society are also attacked. Crenshaw states "… [patriarchy] is a critical issue that negatively affects the lives of not only black women, but of Black men as well" (Crenshaw 1991; p. 1295).

IV. The Advantages and Disadvantages of the Theory of Intersectionality

First, there inherently exists oppression in American society as affected by Western ideology. There are axes, such as race, gender and class that exist in the world and may affect different social institutions. Intersectionality demonstrates that not all experiences are equal within even one society. Women's experiences about the world are not equal, nor are their experiences with other women or men. Intersectionality also illustrates the patterns of wrongness and inequality present in our legal system.

However, intersectionality fails epistemically and this is due to a few realities that intersectionality fails to address. I accept the fact that there actually exists oppression, and that therefore there must be an oppressor and an oppressed in the capitalistic and legal spheres of Western society. Further, I accept that the experiences of women are different and that injustice in society may reflect attitudes of negativity among different individual subjects. However, with regard to using intersectionality as an analytical tool to fight injustice and promote equality, for example, intersectional analysis becomes problematic (Crenshaw 1991; p. 1299; Nash 2008; p. 3). [26]

We may be able to state that there is oppression in any given society but to say that we can know anything about the way in which that oppression governs our personal experiences is faulty. With regard to the concept of 'identity-based politics', the key operative word is 'identity'. This implies some introspective insight into the way in which a woman shapes her identity. Intersectional analysis asserts that we look at the female experience while considering axes of race and gender that are not mutually exclusive, but it becomes crucial to ask in what way are these categories not mutually exclusive on the individual level? There is not enough attention given to the way in which a woman identifies herself and I find this to be an issue in the categorization of the 'white' experience, the 'black' experience and the like.

Further, categorizing 'black women' at all, I argue, is inherently unjust and racist. To give attention to the 'blackness' of an experience is to further marginalize that group to which the blackness applies. By merely qualifying the black experience as something measurable and unitary (Nash 2008; p. 8),[27] Crenshaw is promoting a separate 'black' identity that fundamentally shapes the black female experience, which I argue actually shifts the focus from gender being the primary form of the black female's existence to race being more essential. In actuality, the only semantic difference between the 'black woman's experience' and the 'white woman's experience' are two words both of which are exclusively tied to race. Identifying something as the 'black female experi-

ence' is much more complicated than intersectionality presumes, as there is no such thing. Within a social group different attitudes about their place in the intersection may complicate the way in which Crenshaw stresses the applicability of intersectionality. To categorize women on the basis of race is intrinsically divisive; it defeats the purpose of intersectionality as it is used for theoretical purposes beyond an analysis of the legal system. If one were to consider the ways in which individual black women identified themselves in terms of these axes of oppression, and determine themselves in the intersection of such axes, the result would be so scattered there would be no uniformity. Thus, the 'black female experience' would be a wide array of experiences categorized that are only tied together by race. This does not do anything for the progression of feminism with regards to its potentiality to combat social inequality and injustice. It is absurd to subscribe to the same intrinsically problematic notions that are the subject of critique. To categorize black women singularly is to marginalize the variety of women within the black community and presuppose a connection with that community that may not exist. If intersectionality is meant to include more than just white males and females and is intended to be an actual framework within which we should formally consider our relations with one another, it should be free of the categorical limitations of race.

It is also yet to be determined within the framework of intersectionality whether a black person of Dominican or Cuban descent is still considered black, whether a pacific islander is considered Asian, etc. and whether any of these categories truly mean anything in terms of how we can achieve equality in any meaningful way. Intersectionality incorrectly assumes that we have a basic understanding of (at least) race and gender roles. Whether ethnicity has anything to do with how we categorize people on the basis of race is also unknown, and this is a key missing component to the theory that would either enhance or dismantle its validity.

The lack of attention to the social disadvantages of women of color is apparent in the writings of the most popular Western

feminist philosophers and has been a negative aspect of feminism as a theoretical concept. It is one thing to criticize the unjust social system in which we live, but quite another to assert that we have a uniform way in which we view the social system. It can be argued that Crenshaw herself leaves out other women of color and white women as examples within her framework. This illustrates the problematic issue of the race distinction that Crenshaw fails to make: it marginalizes 'black women' and other women of color[28]. It is important, when formulating a philosophical theory of this nature, [29] to dismantle the stereotypes that corrupt society. By prioritizing the black experience to convey how intersectionality works and saying that other women of color have an experience anything like the way in which Crenshaw describes, intersectionality reveals its insufficiencies as an applicable theory.

In order for intersectionality to serve as a framework that will benefit feminist philosophy, it would have to integrate sociology, psychology and philosophy. It must integrate these other fields in a way that would immunize it against the significant counter argument regarding the difference between the way in which society views individuals on the micro and macro levels and the way in which they identify themselves, both of which inform one another. However, it fails to do this. Socially, it is problematic to assume that we all know the way in which social systems work; that we think about them similarly and that we are all affected in similar ways introspectively. Psychologically, it assumes that we are all mentally affected in terms of social advantage and disadvantage in similar ways among these pre-set unequal groups that Crenshaw redefines[30] for us (black women, white men, etc.); that is, the relationship between the oppressed and oppressor. Philosophically there are too many epistemic holes in this theory, so it becomes difficult to take it to the next level and apply it as a general framework. In order to bridge the gap between sociology and philosophy this element of psychology is necessary.

V. Possible Solutions for the Problems with Intersectionality

Intersectionality needs to be formulated in a way that accounts for the varying nature of the female experience. As it stands, the theory acts contrary to its intentions because it just categorizes the female experience in a different way and places women at various intersections. It is impossible to know where women stand in order to apply the existing framework to determine 'groups'. Intersectionality requires a definition of how axes of oppression shape experiences and whether or not we can group anyone together without being unequal in our separation, to become plausible[31]. Moreover, it does not acknowledge the difference between introspective identity and social identity[32]. As is, intersectionality requires that women accept their 'social standing' and be conscious of it and of the fact that it is universal. However, this is not so. If an African-American woman who is considered black, middle class, and oppressed on different axes in America travels to Sub-Saharan Africa, she would likely be considered European within Sub-Saharan society because personal identities are not fixed and unchanging. Finally, intersectionality needs to bridge the gap between social science and philosophy with psychology; to unify historical social foundational concepts and awareness of these fields with the epistemology of the female experience and the theoretical strength of intersectionality.

VI. Conclusion

Intersectionality, as it pertains to feminist philosophy, attempts to provide a theoretical framework in which we can view the female experience as the intersection of racial and gender oppression.[33] It has some tremendously strong points. Namely, intersectionality has changed feminist philosophy by providing an account of women of color and the ways in which they have a different standpoint with regard to feminist philosophy in general. However, the assumption that we can know anything non-introspectively

about the female experience is flawed. It also emphasizes the gap between sociology and philosophy—both extremely important foundational disciplines in feminist philosophy—and elevates the black female experience to a would-be standard that seemingly disgraces the white female experience of oppression by comparison. This does not advance the progress of feminist philosophy as a discipline but rather leaves it open to philosophical criticism. Therefore, I propose that the inherent flaws of intersectionality be acknowledged and tweaked to accommodate all women equally not by defining their experience but rather by addressing the difference between the way in which women feel about their own experiences and the way women feel about the experiences of other women. Intersectionality can be utilized as a framework to accurately and non-stereotypically philosophize about the female experience, but only by adding to the theory an account of the epistemological and introspective elements to the experience in addition to the social ones.

Notes

1. Which is conveyed by both Crenshaw and Nash; "For intersectional theorists, marginalized subjects have… a particular perspective scholars should consider… when crafting a normative vision of a just society… to fashion a vision of equality" (Nash, p. 3)
2. It is not my personal intention to dismantle intersectionality, but rather to bring up points about the female position specifically that are not being addressed.
3. From the legal standpoint specifically (Crenshaw, Demarginalizing the Intersection) which suggests societal injustice that influences the legal system.
4. "…The use of black women as prototypical intersectional subjects… Crenshaw offers little attention to the ways in which race and gender function as social processes… that mark all black women in similar ways" (Nash, pp. 7).
5. Being that feminist philosophy is a product of social injustice which is a result of other social inequalities, etc.
6. White women dominating the field of feminist theory talking about their own experience does influence the academic world by subliminally stating something like 'this is our actual experience as us females are concerned,

'all of us females experience the same type of oppression on the hands of the patriarchal system that is Western society' etc.

7. Feminism: [n.] 1: The theory of the political, economic and social equality of the sexes; 2: Organized activity on behalf of women's rights and interests (Merriam-Webster, 1895)
8. Excerpt: "…these similarities and differences [between man and woman] must have an influence on morals… it follows that a woman was specifically made to please a man. If a woman is made to please and be subjugated to a man, she ought to make herself pleasing to him rather than to provoke him… when woman complains about the unjust inequalities placed on her by man she is wrong". (*Emile*, Rousseau)
9. i.e., 'All of us women feel this way'.
10. As a developmental theory
11. Meaning taught in formal institutions in a general women's studies survey class and therefore projected in every day society.
12. Carl Linnaeus (a Swedish-born man, commonly regarded as the 'father of taxonomy') systematized the ranking of organisms. Further, in the works of German-born Johann Blumenbach (primarily in the first edition of *"De generis humani varietate nativa liber"*), follow the Linnaean system, but instead categorizes people into five different races based mostly upon appearance and other characteristics—what we deem 'race'.
13. "There appears to be no established agreement on any scientific definition of race" (1996, American Association of Physical Anthropology, p. 569–570)
14. Regarding religion—"Most [blacks] are infidels, having no understanding of Christ, or Gospel" (Fryer, pp. 12), skin color—"Color, accentuated by the term 'Negro' [Spanish for Black] simply became a signifier for the resumed status of slaves" (Sweet, p.9) and difference in lifestyle—"[Negroes] are a people of beastly living, without a God, law, religion or common wealth" (Hakluyt, 167)… "lustful, carefree heathens" (Jones, 1971).
15. Scientific (Blumenbach) and general consensus regarding the way in which Blacks were fundamentally different from White Europeans (see above; 21).
16. In terms of master–slave, dominator–dominated.
17. What I am referring to is "The construction of identity [what is black?] and the system of subordination based on that identity [can blacks and whites sit together on a train?]" (Crenshaw, Mapping the Margins, p. 1298). Whether or not blacks can sit on a train was the court case Plessy v. Ferguson (Crenshaw, p. 1298) was based on and more recently, the Rosa Parks bus incident in 1955 during which time in most of the United States it was still illegal for Blacks and Whites to sit within the same part of a public transport vehicle (Raines, 1977)

18. Combahee River Collective "The Combahee River Collective Statement" (1977). Class in America is tied to economic standing which is in turn dictated by the Western world's primary economic system, capitalism.
19. Sojourner truth's speech given at the Women's Convention in Akron, Ohio in 1851. As a black feminist, she speaks about the distinguishment that society makes when they talk about women and how they treat women of color. "That man over there says that women need to be helped into carriages, and lifted over ditches, and to have the best place everywhere. Nobody ever helps me into carriages, or over mud-puddles, or gives me any best place! And ain't I a woman? Look at me! Look at my arm! I have ploughed and planted, and gathered into barns, and no man could head me! And ain't I a woman? I could work as much and eat as much as a man—when I could get it—and bear the lash as well! And ain't I a woman? I have borne thirteen children, and seen most all sold off to slavery, and when I cried out with my mother's grief, none but Jesus heard me! And ain't I a woman?"
20. "Discrimination, like traffic through an intersection, may flow in one direction, and it may flow in another. If an accident happens in an intersection, it can be caused by cars traveling from any number of directions and, sometimes, from all of them" (p. 149, Demarginalizing the Intersection).
21. "Black women encounter combined race and sex discrimination… the boundaries of sex and race discrimination… are defined respectively by white women's and Black men's experiences" (p. 143, Demarginalizing the Intersection).
22. The combined effects of practices which discriminate on the basis of race, and on the basis of sex (p. 149, Demarginalizing the Intersection).
23. The white woman is more privileged than the black woman, the black man is more privileged than the black female. "The experiential base upon which many feminist insights are grounded is white, theoretical statements drawn from them are over generalized at best, and often wrong" (p. 155, Demarginalizing the Intersection).
24. She provides an example about immigrant women. "Many immigrant women are wholly dependent on their husbands as their link to the world outside their homes" (p. 1248). This fact is then added to her female gender, national identity, etc. to 'equal' that particular female experience. Intersectionality is said to bring this to light.
25. "I argue that Black women are excluded from feminist theory and antiracist policy discourse because both are predicated on a discrete set of experiences that often does not accurately reflect the interaction of race and gender" (Crenshaw, 140).
26. "Intersectionality may provide the means for dealing with other marginalizations as well… race can also be a coalition of straight and gay people of color, and thus serve as a basis for critique of churches and other cultural institutions that reproduce heterosexism" (Crenshaw, 1299).

27. "Marginalized subjects have an epistemic advantage, a particular perspective that scholars should consider, if not adopt, when crafting a normative vision of a just society…. These [intersectional] strategies enable intersectional theorists to draw on the ostensibly unique epistemological position of marginalized subjects to fashion a vision of equality" (Nash, 3).
28. Black women, white women, Asian, etc.
29. By virtue of her using black women as a paradigmatic lived experience.
30. Incorporating themes of sociology, personal identity, psychology and philosophy
31. Instead of liberating females from it.
32. Class, sexual orientation, national origin, physical ability, intellectual capacity, etc.
33. Intersectionality assumes that the way in which society views you is synonymous with the way in which you view yourself. This is simply not true—personal identity can be wholly separate and distinct from the identity you have in society and cannot help significantly.

Works Cited

"2008 American Community Survey 1-Year Estimate." *US Census*. 2000. Web. 1 Apr. 2010.

Blumenbach, Johann F. "De Generis Humani Varietate Nativa Liber." *Johann Friedrich Blumenbach*. University of Minnesota, Mankato. Web. 2010.

Combahee River Collective. "Combahee River Collective Statement." (1977): 1. Web.

Crenshaw, Kimberle. "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics." *University of Chicago, Illinois Legal Review* (1989): 139–67. Print.

Crenshaw, Kimberle. "Mapping the Margins: Intersectionality, Identity Politics, and Violence Against Women of Color." *Stanford Law Review* 43.6 (1991): 1241–299. Print.

DeBeauvoir, Simone. *The Second Sex*. New York: Alfred A. Knopf, 1952. Print.

Hernandez, Victor M. "What Is Race." *American Journal of Physical Anthropology* 101 (1996): 569–70. *American Association of Physical Anthropology*. Web. 04 Apr. 2010. <http://www.culturediversity.org/what%20is%20race.htm>.

Hill Collins, Patricia. "It's All in the Family: Intersections of Gender, Race and Nation." *Hypatia* 13.3 (1998): 62–80. Print.

Nash, Jennifer C. “Re-thinking Intersectionality.” *Feminist Review* 89 (2008): 1–15. Print.

Palmer, Kimberly. “Review: The First Feminist.” *The Women’s Review of Books* 23.4 (2006): 30–31. *JSTOR*. Web. 4 Apr. 2010.

Raines, Howell. My Soul is Rested: Movement Days In the Deep South Remembered. G.P. Putnam Sons. 1977. Print.

Rousseau, Jean-Jacques. “Emile.” *Women, the Family and Freedom: The Debate in Documents (Stanford University Press)* 1 (1983): 43–49. *Liberty, Equality, Fraternity: Exploring the French Revolution*. Web. 07 Mar. 2010. <http://chnm.gmu.edu/revolution/d/470/>.

Sweet, James H. “Collective Degradation: Slavery and the Construction of Race.” *Gilder Lehrman Center International Conference (Yale University)* (2003): 1–33. Print.

Taylor, Barbara. “Mary Wollstonecraft and the Wild Wish of Early Feminism.” *History Workshop Journal* 1.33 (1992): 197–219. *JSTOR*. Web. 4 Apr. 2010.

Truth, Sojourner. “Ain’t I a Woman?” *Modern History Sourcebook* (1851): 1. Web. <http://www.fordham.edu/halsall/mod/sojtruth-woman.html>.

A Defense of Intersectionality

Italia Iossif

Introduction

Intersectionality is a theory rooted in the desire to bring to the foreground people marginalized by the very movements that claim them as their subjects. Historically, both feminist theory and anti-racist politics viewed axes of oppression such as racial oppression and gender oppression in terms of singular or mutually exclusive categories (Crenshaw 1989, p. 139). Not only were these categories mutually exclusive, they take as the content of their meaning the experiences of only a few within those perceived categories. For example, the category "black," while supposedly representing the experience of all black people, truly only represents the experience of black middle class males. The same is true within the feminist movement. The category "woman" presumably represents the experience of all women, but it truly only refers to white middle class women (Crenshaw 1989, p. 139n3). The exclusivity of the categories and their lack of extensive content create problems for women of color within civil rights and feminist movements by rendering their unique experiences of gendered racism invisible. Moreover, such exclusivity also renders women of color invisible within the current legal system, often leaving women of color with little recourse against discrimination focused on them because of their multidimensional identities.[1]

In the late 1980s Kimberle Crenshaw introduced the term "intersectionality" as a metaphor to critique the ways in which a 'single-axis' frame work renders the experiences of women of color invisible within the legal system (Crenshaw 1989, p.139–167; Nash 2008, p 1–15). Intersectionality is the notion that subjectivity is composed of mutually constructed axes of oppres-

sion that interconnect or intersect to form the unique experience of multiply oppressed individuals (Crenshaw 1989, 153–157). Crenshaw explains that, "the aim of intersectionality is to demarginalize those who are multiply-burdened and whose claims of discrimination have been obscured by discrete forms of oppression" (Crenshaw 1989, p. 140). Although Crenshaw was the first to name the concept, intersectional thinking pre-dates Crenshaw, growing "from an already existing discourse of women of color" (Garry 2010, p. 1).

Since its introduction in the late 1980s, intersectionality has been one of the leading tools used by feminist and anti-racist scholars to consider relationships of oppression. Some even hail intersectionality "as one of the most important contributions to feminist scholarship" (Davis 2008, p. 68). Others view intersectionality as the best framework to enable philosophers and feminist theorists to become part of the solution to end oppression (Garry 2010, p. 3). Part of the theory's appeal is that it finally enabled anti-racist and feminine theorists to,

> have a way to discuss the interactions among race, ethnicity, gender, sexual orientation, social class, and so forth, that : i) rejected ranking them, and ii) enabled both… activism and… theories to move forward with at least a minimal understanding of the ways in which the major axes of oppression interacted with each other and affected human lives, and helped [to] start to grasp the complexities of [the] similarities, differences, and the networks of hierarchical relationships among us [feminist /anti-racist theorists as well as people in general] (Garry 2010, p. 3).

Although well-received by many within feminist discourse, there are still those who find difficulties with the theory of intersectionality.

In 2008, Jennifer Nash published her critique, "Re-thinking intersectionality." In this essay, Nash argues that intersectionality, as a view, is incomplete because it lacks a clear methodology. Furthermore, she argues that the use of black women as the quint-

essential subjects for intersectional analysis has made unclear who is intersectional. That is, one must ask whether intersectionality is a general theory of identity or a theory of how multiply marginalized groups form identities. In this paper, I will closely examine the arguments Nash presents in favor of these two criticisms of intersectionality. I will argue that a closer reading of both of Crenshaw's seminal essays ("Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Anti-discrimination Doctrine Feminist theory and Antiracist Politics" and "Mapping the Margins: Intersectionality, Identity Politics, and Violence Against Women of Color,") reveals that many of Nash's criticisms can be answered through Crenshaw's own analysis. Furthermore, I will suggest that intersectionality has been successful precisely because it is vague enough to be used by feminists as a tool that is useful across a range of fields.

A Lack of Methodology

Nash's first criticism of intersectionality is that it lacks a clear methodology, a problem with both theoretical and practical effects (Nash 2008, p. 4–6).[2] To be more precise, by "theoretical" she does not mean strictly that there are theoretical consequences of holding an intersectional view. What she means is that, because intersectionality lacks a clear methodology there are different "theories" of or approaches to how to perform an intersectional analysis (Nash 2008, p. 4–6). She explains, "little has been said on how to study intersectionality, that is, of its methodology" (McCall 2005, p. 1771; Nash 2008, p. 4). Yet, Crenshaw's description of intersectionality as a concept (1) centers on multiply oppressed individuals, (2) allows for the description of concurrent "multiple oppressions," and (3) provides an understanding of the complexity of identity by "politically progressive projects." This seems to suggest that at the very least Crenshaw has some guiding principles as to what intersectionality ought to do (Crenshaw1989, p. 140; Nash 2008, p. 4). Nash also claims that intersectionality has no way of examining the lives of "multiple subject positions

(Nash 2008, p. 5)."

Despite descriptions of methodology like the ones above, feminists such as Leslie McCall have attempted to create methodological tools to help us deal with the all the intersections that constitute identity. McCall's attempt to add to the methodological approach of intersectionality is an example (for Nash) that intersectionality as it stands is not equipped to handle the myriad of combinations that constitute identity and that intersectionality is not a complete theory. However, I would argue that Crenshaw herself shows us the ways to deal with the multiple intersections that constitute identity and therefore such attempts are unnecessary

Before I continue to provide evidence of how Crenshaw herself shows us the ways to use intersectionality, I want to make clear what I am arguing and what is at stake in this discussion. I have two kinds of views working here. The first is that I believe that if we find a problem with a theory or concept, it perhaps is best to return to the person who originated the theory to see if in fact there is a problem or to determine if the problem is simply with those employing it. In this case, I am trying to prove that perhaps the problem is not the lack of a clear or explicit methodology as Nash says it is. Rather, that we just need to pay closer attention to how Crenshaw is using intersectionality to ensure that we are following the proper guidelines in the performance of our intersectional analyses. However, I do not believe it is always a problem for theories to be used in ways or in areas that the author did not originally intend them to be used. For example, many ethical theories have components of economic theories that support them. What needs to be determined then is if the use of a concept/theory in different ways is incorrect in light of the spirit of the theory.[3]

"Mapping the Margins" she discusses two forms of intersectionality. One form is "structural intersectionality" (Crenshaw 1991, p. 2). Structural intersectionality explains how Crenshaw intends intersectionality to function as a description of identity. Crenshaw describes structural intersectionality as a way to determine "the ways in which the location of women of color at the

intersection of race and gender makes our… experience(s)… qualitatively different than that of white women" (Crenshaw 1991, pp. 2–5).[4] Granted, focusing here on race and gender, she does not account for all the other possible intersections. Yet we can extrapolate that we can continue to add categories and the methodological approach of the analysis would remain the same. In other words, Crenshaw could have just as easily said, "structural intersectionality is a way to determine the ways in which the location of women at the intersections of race, gender, class, heterosexism, able-bodiedness, etc… makes experience qualitatively different than that of white women." Nash argues that Crenshaw's lack of attention to those categories suggests a problem with intersectionality (Nash 2008, p. 7). However, there are two very simple responses to Nash's objection. First, it is possible that Crenshaw did not have the time or space to go into all of those possibilities and describe how such a position is qualitatively different, nevertheless we can still imagine that there is such a location as imagined above.[5] Second, Crenshaw does talk about issues beyond race; she discusses how, in the U.S., immigrant women's experiences of sexual violence are constructed in relation to English as dominant language (Crenshaw1992, pp. 3–10). Moreover, this seems to suggest that by women of color she is not limiting her analysis to black women.

The second form of intersectionality that Crenshaw describes is called "political intersectionality" (Crenshaw 1991, pp. 2, 5–12). Political intersectionality, as Crenshaw describes it, deals with the ways in which women of color have been marginalized by antiracist and feminist movements. For instance, political intersectionality explains how these movements have not truly been able to tackle the problem of violence against women, because they fail to recognize how violence functions in the lives of women of color" (Crenshaw 1992, pp. 5–14). This form of intersectional analysis enables us to see how women of color are marginalized within political movements. Not only is intersectionality suited to handle multiple intersections, but it is also able to critique movements by pointing out that their agendas are not built around the

experiences of everyone they claim to represent.

Thus far, I have described the purported theoretical problems that Nash believes occur because of intersectionality's lack of a clear and exact methodology. I hope to have shown that, despite a purported lack of methodology intersectionality as Crenshaw describes it enables us not only to understand identity but also provides a way to handle the countless intersections that constitute identity. Intersectionality also provides a means to criticize feminist projects or anti-racist projects that limit themselves to monistic views of oppression, or provide minimal mitigation by failing to recognize intersectionality. Although I believe I have dismissed the "theoretical" problems that arise because of a lack of methodology there still remain what I call the practical problems.

Practical problems of intersectionality have to do with how intersectional analyses look when employed. Nash argues that although intersectional "practices work to disrupt cumulative approaches to identity, and to problematize social categorization through strategic deployments of marginalized subjects' experiences, intersectional projects often replicate precisely the approaches they critique" (Nash 2008, p. 6). In other words, many intersectional projects commit the error of using the additive approach. The cumulative approach is the view that identity is created qua black + lesbian + woman + low socioeconomic status = identity. Although intuitively we may think this is true, the problem with this view is that the categories used do not represent the universal claims that they purport to represent. So with the additive approach what really happens is that we are saying black man + white woman + lesbian = woman of color. Nash discerns the cumulative approach "in Crenshaw's critique of anti-discrimination law" (Nash 2008, p. 6). Crenshaw states:

> Because the intersectional experience is greater than the sum of racism and sexism, any analysis that does not take intersectionality into account cannot sufficiently address the particular manner in which Black women are subordinated. (Crenshaw 1989, p. 140)

Nash explains that, "for Crenshaw, it is the fixed mutually exclusive categorizations of race and gender that marginalize black women" (Nash 2008, p. 6). Crenshaw is explicitly stating here that identity is more than just the sum of the categories. Therefore, we need to account for how the experiences of women of color are shaped by all the factors before we try to seek to remedy oppression. Additionally, for Crenshaw, the categories are not mutually exclusive. After all, the problem with the monistic view she criticizes is that they view the categories as mutually exclusive (Crenshaw 1989, pp. 140–167).

Crenshaw takes great pains to express the notion that identity is more than the sum of race and gender. However, Nash argues that, the fact that Crenshaw ignores an examination of multiple burdens (e.g. class, sexuality, etc..), and focuses only on the sites of race and gender "renders black women's [experience] the aggregate of race and gender" (Nash 2008, p. 7). While it is true that Crenshaw focuses mainly on race and gender in "Demarginalizing" it is not the case in "Mapping the Margins." In "Mapping the Margins" Crenshaw specifically addresses class, and immigrant status of women (Crenshaw 1992, p. 310). Therefore, she is not limiting the description of women of color to race and gender and rendering them the aggregate of such. Moreover, we cannot ignore that her aim in "Demarginalizing" is narrow; she is looking only at a particular intersection and seeing how it affects those women legally and in the work place. Finally, regardless of the fact that Crenshaw only engages race and gender in her discussion about race and gender one cannot ignore her claim that "intersectionality is greater than the sum of racism and sexism" (Crensahw1989, p.140)!

It is true that we tend to speak in additive terms. For example, I might say that "in order to understand how oppression functions in the life of a multiply marginalized subject we need to see how the categories are enmeshed or come together, etc." However, in saying this, I am merely limited by language. It is difficult to express the concept of intersectionality through categories converging without sounding like I am adding catego-

ries together. Nevertheless, Crenshaw is explicit in saying that she recognizes the problems with the existing categories (Crenshaw 1989, p. 139n3). Therefore, it is not clear to me that Crenshaw is committing the cumulative approach error.

Nash's final criticism of Crenshaw is that she does not pay attention to the fact that race and gender have worked in "distinctive ways for particular black women throughout history. The use of black women's experience as Crenshaw has described it seems to suggest that in the identities of black women, race and gender are… trans historical constants that mark all black women in similar ways" (Nash 2008, p. 7). This reading of Crenshaw strikes me as incorrect. At the very least, it fails to recognize that Crenshaw is using intersectionality to perform a legal analysis in *Demarginalizing*. Discrimination is conceived as occurring at a fixed point in time, and so in legal analyses of discrimination, history is not methodologically important. More to the point, the law in the cases Crenshaw describes was not concerned with history, either, in its determination of discrimination. The court just wanted to establish if discrimination had occurred at a given point in time (Carastathis, n.d.; Crenshaw, 1989, pp. 139–167). Furthermore, the analysis Crenshaw performs in *Mapping the Margins* on violence against women is clearly set in contemporary times, thus suggesting that such is the current state of affairs and that the analysis is applicable to this historical moment. Furthermore, intersectional analyses often aim to show how contemporary institutions fail to recognize the intersectionality of victims and therefore fail to mitigate the painful circumstances they were out in place to mitigate. An example is the battered women's shelter whose mission is to help all women, but fails to do so because it does not accept non-English speaking victims (Crenshaw 1992, pp. 3–7). Therefore, it is clear that Crenshaw's intersectional analysis is limited to a current time and place, the U.S. in the 1980s and 1990s.

Thus far, I have presented three problems with intersectionality that Nash claims arise due to its lack of a clear methodology. The first two were theoretical and the third was a problem of

praxis. My aim was to show that none of these supposed problems were actual problems. My intention is to prove that, although there is no enumerated methodology on how to perform an intersectional analysis, there are at least two forms of intersectionality described by Crenshaw that enable it to be diverse and there is at least a definition which is useful as a guideline. I want to suggest that the problem is not that intersectionality lacks methodology but, rather, it offers loose guidelines. For example: we ought to put women of color at the center of our analysis, however we are not limited to what we can analyze, we can analyze anything from feminist philosophy to legal theory and practice. Even colonialism can be analyzed using an intersectional approach. Precisely because the theory is *not* too narrow or demanding that theorists in all fields (e.g., philosophy, law, sociology) are enabled to seriously consider whether their work as feminists truly aims to end oppression for all. It is for this reason intersectionality has been so widely accepted.

> Intersectionality offers endless opportunity in interrogating one's own blind spots and transforming them into analytic resources for further critical analysis. In short, intersectionality, by virtue of its vagueness and inherent open-endedness, initiates a process of discovery, which not only is potentially interminable, but promises to yield new and more comprehensive and reflexively critical insights (Davis 2008, p. 78).

Kathy Davis argues that the infinite number of possible combinations of categories and lack of definition have made it possible for endless "constellations of intersecting lines and differences to be explored" (Davis 2008, p. 78). Part of the appeal of the theory is that we are able to draw out the possibility of another axis of oppression, thus continuing to further the aim that no one be marginalized (Davis 2008, p. 78). Furthermore, it seems that what is truly important is that we should be able to address the pressing issues of marginalization and of oppression in such a way that white women and black men are not the only ones benefiting from

the solutions offered by single-axis political movements.

Who Is Intersectional?

As stated earlier, the use of black women as paradigm intersectional subjects has raised the question, who is intersectional? Is it only marginalized people, or are people who are privileged on some axes intersectional as well? In this section, I will examine what constitutes an intersectional being and what kind of identity is important for the use of an intersectional analysis.

In other words, Nash's criticism of intersectionality is that, due to the use of black women's experience as a common starting point of intersectional analysis, it has become difficult to determine what exactly intersectionality purports to do. Moreover, how to go about using intersectionality is also unclear because of the vague definitions. In this section I will argue that it is not the case that the use of the "black woman's experience has obscured the question of whether all identities are intersectional or only multiply marginalized subjects have an intersectional identity" (Nash 2008, p.10).

In "Demarginalizing the Intersection," Crenshaw uses the metaphor of an intersection to explain how she understands intersectionality. "Discrimination" she explains, "like traffic may flow in one direction and it may flow in another. If an accident happens in an intersection, it can be caused by cars traveling from any number of directions and, sometimes, from all of them (Crenshaw 1989, p. 149). This is important because it reveals that Crenshaw understands intersectionality to be a way of understanding an individual's location at the intersection of multiple discriminations as well as privileges. Although she does not explicitly say privileges because she is talking about discrimination, arguably identity categories exist for even the privileged. All people are intersectional by virtue of the fact that they are a part of multiple identity categories. For example "white" is a race category, "gay" is a sexual orientation category, and "male" is a gender category, none of which excludes any of the other. Furthermore, in "*Mapping*

the Margins," she explains structural intersectionality as a way to understand individual locations. She states, "My objective here is to advance the telling of that location by exploring the race and gender dimensions of violence against women of color…. [S]tructural intersectionality, [describes] the ways in which the location of women of color at the intersection of race and gender makes our actual experience of domestic violence" (Crenshaw 1991, pp. 1–2). On Crenshaw's view, all people find themselves at the center of some intersection, whereas only *some* people may experience a collision of multiple "flows" of discrimination. On this understanding of intersectionality, it does not seem to be the case that the use of the black experience convolutes who occupies an intersectional location or identity (Nash 2008, p. 10). Intuitively, it is not the case that only multiply marginalized individuals have intersectional identities.[6] The use of the experiences of women of color at the center of the intersection of race and gender is merely a way to show that monistic understandings of the categories of race, class, and gender, neglect the overlapping relationship between them. Therefore, the distinction Nash makes between all identities being intersectional or only marginalized identities being intersectional is not applicable.

It seems to be the case that intersectionality can be viewed as a kind of tool that enables an understanding of the particular or unique position of multiply marginalized individuals. In this case, however, Nash argues that intersectionality theorists need to create a theory of identity for understanding how "race, gender, sex, and class, among other categories, are produced through each other, securing both privilege and oppression simultaneously" (Nash 2008, p. 10). My question for Nash then is whether a theory of identity would account for personal identity or for how others view us (social identity). A theory of the way(s) in which personal identities are formed in this case may not be helpful because individuals are targeted for discrimination on the basis of their outward appearance to others, not on their subjective belief about who they are. Intersectionality helps us to see how discrimination looks when imposed on an individual on the basis of multiple

identities. In short, there is no need for a theory of subjective identity. As stated earlier, intersectionality already helps us to find identity through location. In other words, we are able to determine a kind of social identity by considering the intersections. What is needed, or what seems crucial to all intersectional theorists, is an understanding of the identity categories of "race", "sex", etc and how those when narrowly understood, are oppressive.

Conclusion

Intersectionality can be viewed as either a tool or a concept that aids us in the demarginalization of those who are multiply oppressed and whose experiences are ignored because of monistic views of oppression. Despite several strong charges that intersectionality fails because of vagueness and an inability to determine who counts as intersectional, a return to the originator of the term—Kimberlé Crenshaw—reveals that intersectionality is fully capable of doing that for which it was created. In essence, it is able to make progress and force theorists out of their shell and possibly to reexamine their own racist/sexist perspectives. Finally, intersectionality is able to move people. It is successful because it is not so narrow that it cannot deal with the multiplicity of its variously located subjects. In short, precisely because it's vague, it enables us to use it in a variety of ways to render visible a multiplicity of subjective locations.

Notes

1. In "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist theory and Anti-racist Politics" Crenshaw focuses on the 3 court cases. They are: *DeGraffenried v. General Motors, Moore v. Hughes Helicopter, and Payne v. Travenol* . These cases illustrate how the single-axes framework was not sufficient enough a frame work to handle the multidimensional identities of black women.
2. In her essay Nash splits up her criticisms into two parts "theories of intersectional methodology" and "practices of intersectional methodology" and so here I have kind of taken on Nash's distinction by explaining the problems of a lack of methodology as resulting in both "theoretical" and "practical" problems (Nash, 4–6)

3. While McCall makes a similar criticism of intersectionality, it is important to note that she is a sociologist and Crenshaw is a law scholar so it is possible that her need to make adjustments to the methodology of intersectionality stem from the fact that she is working in a different area. The question then is, are her additions consistent with the view or the intention of intersectionality as represented by its originator? But this is a question that requires a through reading of McCalls paper and that is outside the scope of this paper.
4. So, I can see here a potential problem. Crenshaw references white women; however, we need to recall the context of this article it is about the black experience but that does not mean it is limited to that, again she could have easily been a disabled person making the claim that her experience was qualitatively different than that of an able bodied person. I think it is a mistake to limit or criticize Crenshaw for speaking only about women of color when her goal is to demarginalize them. It is possible that that is a mistake on her part; however, that does not mean intersectionality is limited in the same way.
5. It is common for philosophers to bring up a topic and willingly admit that while the subject is important it is not within their capacity or the scope of the paper to discuss that topic.
6. Multiply marginalized cases are not the only ones with intersectional identities. This might not be clear because all the discussions involve only the experiences. The reason is simple, it is unnecessary to bring non-marginalized identities to the center of discussion, because they have been there fore years. Furthermore, there is an underlying belief that if the most marginalized are freed the rest will come too. So much of intersectional work aims at freeing those who it views as most marginalized. It is debatable whether African American women are the most marginalized, considering native American plight. Never-the-less bringing the experience of native Americans center stage would not conflict in anyway with intersectional agenda I don't think.

Bibliography

Chang, R. S., and J. M. Culp. "After Intersectionality." *University of Missouri City Law Review* 71 (2002): 485–91.

Crenshaw, Kimberle. "Demarginalizing the Intersections of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory, and Antiracist Politics." *University of Chicago Legal Forum* (1989): 139–67.

Crenshaw, Kimberle. "Mapping the Margins: Intersectionality, Identity Politics, and Violence Against Women of Color." *Stanford Review* 43.6 (1991): 1241–299.

Davis, Kathy. “Intersectionality as a Buzzword: A Sociology of Science Perspective on What Makes Feminist Theory Successful.” *Feminist Theory* 9 (2008): 67–86.

Garry, Ann. “Intersectionality, Metaphors and the Multiplicity of Gender.” 1–50. forthcoming.

McCall, L. “The Complexity of Intersectionality.” *Signs* 30.3 (2005): 1771–800.

Nash, Jennifer. “Re-thinking Intersectionality.” *Feminist Review* 89 (2008): 1–15. <www.feministreview.com>

Psychopaths and Moral Rationalism

Carlos Maguina

I. Introduction

The use of psychopaths in ethics is part of a long tradition of invoking colorful, if usually fictional, entities in philosophical discussions. Philosophers have used Black-and-White Mary, spectrum inverted people, doppelgangers, zombies, brains in a vat, Schrodinger's cat, Wittgenstein's beetle, and a host of other colorful entities, as a means to illustrate inconsistencies and emphasize certain commitments, whether philosophical or pre-philosophical. In short, these colorful entities are theoretical tools intended to help us see things that are easily overlooked. In what follows, we'll consider Shaun Nichols' use of psychopaths as an attempt to illustrate that emotions play a role in forming moral judgments and, ultimately, as a way to undermine Empirical Rationalism, a version of Moral Rationalism which states that moral judgments are rational—not emotional—judgments.

In "How Psychopaths Threaten Moral Rationalism" (Nichols 2002), Nichols argues that our conception of psychopaths renders Conceptual Rationalism false, and that the empirical evidence on psychopaths renders Empirical Rationalism implausible. It is the latter claim that I take up here. I will argue, pace Nichols, for the plausibility of Empirical Rationalism.

II. Nichols and Empirical Rationalism

Nichols sees Empirical Rationalism as, in part, attempting to secure "a kind of objectivism about moral claims (Nichols 2002, p. 1). He argues that if Empirical Rationalism were right, it would follow that "creatures who have all the rational faculties that we

do (including aliens) should arrive at the same moral views that we do" (Nichols 2002, p. 11). In other words, if there were creatures who shared our rational faculties but not our moral views, they would serve as counter-examples to Empirical Rationalism. Nichols stops short of claiming that psychopaths are such creatures; instead, he argues for the plausibility that they are. Nichols is claiming that it's plausible that psychopaths share our rational faculties but not our moral views, which, if right, would make psychopaths successful counter-examples to Empirical Rationalism. Let's consider how Nichols arrives at this conclusion.

Nichols makes use of empirical evidence concerning psychopaths' inability to distinguish between moral violations and conventional violations. Moral violations—e.g. pulling hair, pushing, hitting, etc.—are construed as emotion-eliciting situations. Conventional violations—e.g. talking out of turn, playing with one's food, chewing gum in class, etc.—are construed as non emotion-eliciting situations. The empirical evidence shows that children (and "a host of others" such as autistic individuals) do distinguish "the cases of moral violations from the conventional violations on a number of dimension," most notably permissibility, seriousness, and justification (Nichols 2002, pp. 13–14). For example, when asked to justify why hitting is wrong (a moral violation), children answer in terms of the harm to victims (a moral-type justification); psychopaths, on the other hand, respond in terms of social acceptability (a conventional-type justification). According to Nichols, what accounts for this difference is "a difference in affective response" (p. 20). That is, it's children's emotive responses in considering harming others that explains their ability to distinguish moral violations from conventional violations. The presumption is that psychopaths are at least as rational as children. This is important for Nichols because, otherwise, Empirical Rationalists could maintain that children are drawing on their superior rational capacities in order to make the moral/conventional distinction—which would undermine Nichols' argument. Anticipating such a response, Nichols points out that it is a putative fact that psychopaths are at least as rationally capable as children (and

the host of others), so it doesn't seem likely that the psychopath's inability to make the moral/conventional distinction is on account of a rational deficiency on his part. Once it is ruled out that rational capacities can play an explanatory role, we can safely infer that it is superior emotional capacity of children that enables them make the moral/conventional distinction that psychopaths fail to make. Nichols characterizes the challenge facing Empirical Rationalists as follows:

> the Empirical Rationalist needs to find a rational defect in psychopaths that explains their deficit in moral judgment; and this deficit should not be present in autistic individuals, young children, control criminals, and a host of other rationally idiosyncratic humans who don't share the psychopaths' deficit in moral judgment (Nichols 2002, p. 20).

As Nichols sees it, the implausibility of successfully meeting this challenge makes Empirical Rationalism implausible. Let's consider his argument in detail.

According to Nichols, if moral judgments were rational judgments, then psychopaths as well as children would be able to make the moral/conventional distinction because psychopaths are at least as rationally capable as children. But psychopaths aren't able to make the moral/conventional distinction, so moral judgments must not be rational judgments; instead, they must be emotional judgments. The force of Nichols' argument rests in its simplicity:

> Premise 1. Psychopaths are emotionally deficient (empirical fact)
>
> Premise 2. Psychopaths are at least as rationally capable as children (strong intuition)
>
> Premise 3. Children make the moral/conventional distinction (empirical fact)
>
> Premise 4. Psychopaths don't make the moral/conventional distinction (empirical fact)

Premise 5. Only rational and emotional capacities can play an explanatory role in explaining how one can make, or fail to make, the moral/conventional distinction (tacit and plausible assumption)

Premise 6. This difference in ability between children and psychopaths isn't a matter of rational ability (plausible inference from P1–P4)

Premise 7. This difference in ability between children and psychopaths is a matter of emotional ability (inference from P5–P6)

Premise 8. Empirical Rationalism states that moral judgments are rational judgments (statement of fact)

Conclusion: Empirical Rationalism is false (follows from Premise 7 and Premise 8)

The above argument relies on a neat dichotomy between moral violations and conventional violations in which the former are construed as emotion-eliciting situations and the latter as non emotion-eliciting situations. In addition, moral violations elicit emotions because these situations involve hitting someone, pushing someone, pulling someone's hair, etc—they have a 'harming others' aspect to them. Conversely, conventional violations fail to arouse the emotions because there is no readily identifiable victim to suffer harm in these situations—they lack the aspect of harming others. But this neat picture gets disrupted once we introduce a third type of violation, moral impersonal-violations, construed as non-emotion-eliciting situations which have a 'harming others' aspect to them. Before we wield 'moral impersonal violations' against Nichols' position, let's make sure we're on solid empirical footing.

III. Moral-Impersonal Violations

In an fMRI study, Joshua Greene (Greene et al. 2001, p. 2105–7) monitored the activity of areas in the brain associated with

emotion and areas associated with working memory in subjects who were asked to consider moral-personal, moral-impersonal, and non-moral dilemmas (i.e. conventional violations). The tests showed that there is little activity in areas of the brain associated with emotion (measured in percent change in MRI signal relative to a baseline) in subjects considering moral-impersonal dilemmas, dilemmas in which there is an identifiable victim to suffer harm but in which no intimate interaction is required between subject and victim. In other words, these are the third type of violations I mentioned: non-emotion-eliciting situations that have a 'harming others' aspect to them. The paradigm case of moral-impersonal dilemmas is the Trolley Car dilemma: a dilemma in which we have the option of either letting a train kill five people, or pulling a lever that will sidetrack the train, thereby saving the five people but killing one individual on the sidetrack. The non-intimate interaction is the pulling of the lever. In contrast, moral-personal dilemmas are just like moral-impersonal dilemmas except they involve intimate interaction between subject and victim. The paradigm case is the Footbridge dilemma: a dilemma in which we have the option of either letting a train kill five people, or pushing a large stranger from a footbridge onto the tracks, thereby killing him and saving the five people. What we should take away is that moral-impersonal violations, while having a 'harming others' aspect to them, do not elicit emotions because they do not require intimate interaction between subject and victim. I will utilize moral-impersonal violations to highlight several problems with Nichols' analysis, and argue that these problems undermine his attack on Empirical Rationalism.

IV. Defending Empirical Rationalism

Nichols characterizes the moral/conventional distinction as a distinction between an emotion-eliciting situation and a non-emotion-eliciting situation. The rationale for this is that moral violations have a 'harming others' aspect, while conventional violations do not. But this account is one an Empirical Rationalist

can embrace, as he will want to say that, children are able to make the moral-impersonal conventional distinction precisely because moral-impersonal violations have a 'harming others' aspect that conventional violations do not. I am unaware of any empirical tests done on the matter, but this certainly falls in line with our intuitions about the capacities of children. It's difficult to believe that children would somehow fail to make this distinction given the 'harming others' aspect of moral-impersonal violations and the lack thereof in conventional violations. For example, consider the following dilemma: you have the option of letting a train continue on its track towards your stalled car, or pulling a lever that redirects the train towards a man on the sidetrack. As to the question whether children could make the oral-impersonal/conventional distinction, it's not open to Nichols to say that they could. This is because the emotions, which Nichols must invoke as playing an explanatory role, aren't aroused in either type of violation. If he accepted that children could make the distinction, then, in light of Greene's fMRI study, he would have to accept that it's a distinction they make strictly on rational grounds. The most plausible explanation would be that moral-impersonal violations have a harming others aspect and conventional violations do not. The difference now, when we give this analysis about the moral-impersonal/conventional distinction, and before, when Nichols gave the same analysis about the moral-personal/conventional distinction, is that we know that the 'harming others' aspect alone doesn't arouse the emotions, and it is the emotions that Nichols must ultimately invoke as playing an explanatory role. The problem for Nichols is that it seems plausible that children could make the moral-impersonal/conventional distinction, and thus that their ability to make the moral/conventional distinction is fully explained by the fact that moral violations have a 'harming others' aspect and conventional violations do not. In this case, however, the emotions, or their arousal, are stripped of any explanatory role in how children make the moral-personal/conventional distinction, as this can be fully explained by the 'harming others' aspect of the former and the absence of such in the latter.

Shaun Nichols highlights a puzzling fact about psychopaths: they do not make the moral-personal/conventional distinction. It's puzzling because our intuition is that rational beings should easily be able to distinguish between violations in which innocent people are killed and violations in which you talk with your mouth full. The difference can be put in various ways, perhaps no one precise way, but that those violations that have a 'harming others' aspect are moral violations and those that don't aren't seems precise enough. We're left wondering, then, why psychopaths give conventional-type justifications when presented with moral dilemmas. But others, and this includes adults as well as children, have no problem making the moral-impersonal/conventional dilemma despite the fact that the emotions aren't elicited by either type of dilemma. In fact, it would be surprising to find out that we need our emotions to be aroused in order to make this distinction. We think we can make it on rational grounds, and in fact we do so. But, according to Nichols, despite the fact that moral-personal violations, like moral-impersonal violations, have a 'harming others' aspect, when we as adults make the moral-personal/conventional distinction, we're not making it on the basis that the former has a 'harming others' aspect and the latter does not—that is, on strictly rational grounds. Instead, Nichols will say that the emotions are needed to make this distinction. But this isn't anything we're compelled to accept by anything Nichols has put forth.

V. Conclusion

Empirical Rationalism is forceful because it coincides with the strong intuition that we reason to ethical judgments. We make the moral-impersonal/conventional distinction on rational grounds, which is something Nichols cannot dispute, and what's available for consideration when making this distinction is also available when making the moral-impersonal/conventional distinction, yet Nichols has argued that such evidence isn't what explains our ability to make the moral-personal/conventional distinction. While

psychopaths prove a puzzling case, Nichols hasn't made the case for the implausibility of Empirical Rationalism just because we have a compelling argument that purely rational grounds are sufficient to make this distinction.

Bibliography

Greene, Joshua D., Sommerville, R. Brian, Nystrom, Leigh E., Darley, John M., Cohen, Johnathan D. (2001) "An fMRI Investigation of Emotional Engagement in Moral Judgment," *Science,* Vol. 293, pp. 2105–2108.

Nichols, Shaun. (2002) "How Psychopaths Threaten Moral Rationalism, or Is it Irrational to Be Amoral?" *The Monist,* 85, Issue 2, pp. 285–303.

Kantian Deontology and the Moral Exclusion of Animals

Raj Shah

Introduction

The philosophical issue in bioethics of what moral status non-human animals occupy in the world is one that has great practical relevance in modern society. Indeed, the fact that non-human animals (interchangeably referred to here by the common term, "animals") today are used in the production of a variety of goods means that animals play a very large role in our own lives and in society. Thus, any discussion on what moral worth animals possess and how they should be treated inevitably concerns many such different aspects and industries of society.

Throughout the history of philosophy, different philosophical schools of thought on ethics have argued that animals possess no intrinsic moral worth at all and do not merit any moral consideration or concern. Typically, according to such moral theories, humans possess a certain characteristic that animals lack. For this reason, the differentiation between animals and humans is such that humans can claim justification for treating animals instrumentally. In short, in such theories, all animals are in some way distinguished from all humans and placed into a completely separate moral category of beings. In essence, such theories *morally exclude* animals, that is, they exclude animals from the moral category of beings that actually have intrinsic worth, by stripping animals of moral importance.

However, attempts to make moral distinctions between all animals and all humans are doomed to end in failure. This is mainly for the reason that, as modern biological discoveries reveal, it is simply not possible, in a clear way, morally to differentiate between

humans and other species of living things because of the biological similarities between them and the similar biological origins surrounding all organisms. Attempts by philosophers to isolate a characteristic that makes humans and *only* humans morally valuable are undermined by the fact that animals are biologically capable of attaining such characteristics. Such biological similarities and capabilities pose a challenge to any attempt to build a strict, impassable moral barrier between humans and animals. Consequently, any attempt morally to distinguish between all humans and all animals ultimately seems to breaks down.

Summary

Of course, it is impossible to prove here that *all* arguments for the moral exclusion of animals are wrong. In the end, there are simply too many arguments that advocate the moral exclusion of animals. Thus, in this paper, I will attempt to narrow the focus of criticisms of such theories by examining one specific moral system widely renowned for its moral exclusion of animals: Kantian Deontology, a duty-based conception of ethics. For Kant, the capacity to use rationality is what matters in determining whether a being has moral status. According to Kantian ethics, if an entity is not rational and is unable to comprehend moral principles, then it has no moral value or importance. From this, Kantians have often moved to argue that animals, being supposedly non-rational, have no moral worth.

In this paper, I will attempt to show that the Kantian moral view of animals ultimately breaks down upon analysis and cannot establish an effective moral distinction between humans and animals. In the end, the problem with the Kantian view (and perhaps any theory arguing for the moral exclusion of animals) is that modern scientific developments and biological discoveries are simply not compatible with the notion that there exists a solid and tenable distinction between humans and animals. I will attempt to demonstrate this on two different levels. I will first argue that the Kantian view of the moral status of animals

is problematic because it incorrectly assumes that rationality is something exclusive to human beings and something unattainable by animals. Second, I will examine how Kantian Deontology's insistence on rationality as the sole criterion of moral worth actually ends up morally excluding certain categories of human beings along *with* animals. Ultimately, as I will show, Kantian attempts to solve this second problem defeats the strict Kantian moral distinction between all humans and all animals.

The Kantian View of Animals

Considered a major system of morality today, Kantian Deontology originates from the writings of eighteenth-century Enlightenment philosopher Immanuel Kant. Of course, since then, many later authors have added newer forms of argumentation to Kant's original theories. Nevertheless, at the core of Kantian Deontological theory is the inherent moral value of rationality. Kant, in his great philosophical work *The Groundwork for the Metaphysics of Morals*, established a basis for morality through the rationality of human beings. As Kant states in the Groundwork, "all [that] man is subject to are laws—universal laws—legislated by himself [by means of reason]" (Kant, 1785, p. 31). In other words, for Kant, all human beings are bound by a universal moral law, which they discover for themselves and act on by means of rational analysis. As Kant states, rationality is "designed by nature to be a giver of universal law" (Kant 1785, p. 31).

Yet, how does reason ultimately lead one to formulate and act on universal moral laws? For Kant, human beings discover moral laws for themselves based on whether they would be rationally universalizable, that is, if they could be applied to all other rational beings without contradiction (Kant, 1785, pp. 25-26). If one rationally judges a law as universalizable and applicable to all other rational beings, Kant argues, one should adopt it as an objective moral principle. For example, a person would see that the principle that one should lie to protect one's self-interest would be morally wrong because one would rationally deduce that, if this

principle were universalized, the very purpose of language itself would be self-defeating. From this, one would be able to arrive at the objective moral principle that one should not lie. Thus, because of such rational judgment with which humans are able to distinguish right and wrong, Kant went so far as to describe human beings as "legislators" of universal moral laws (Kant 1785, p. 31). In this way, reason, for Kant, was paramount to legislating and acting on moral laws.

For Kant, then, reason is critical for any moral dignity and value because it is through reason that human beings are able to understand and self-legislate laws of morality. As Kant states, "I maintain that man—and in general every rational being—exists as an *end* in himself and *not merely as a means* to be used by this or that will at its discretion" (Kant 1785, p. 28). In short, for Kant, rational beings have absolute moral value in such a way that they are "an object of respect, and something that sets limits to what anyone can choose to do" (Kant 1785, p. 29). As Kantian philosopher Carl Cohen puts it, human beings' "inherent value gives them moral dignity, a unique role in the moral world, as agents having the capacity to act morally and make moral judgments" (Cohen 1997, p. 773). In the Kantian sense, then, only human beings have any kind of moral value or dignity.

In fact, this view of moral dignity plays a key role in Kantian morality. At one point in the Groundwork, Kant formulates morality in terms of the rule, "Act in such a way as to treat humanity, whether in your own person or in that of anyone else, always as an end and never merely as a means" (Kant 1785, p. 29). Thus, for Kant, humans have an inherent moral dignity that imposes moral duties and obligations on other rational beings. Human beings, according to Kantian moral theory, have the moral duty always to treat one another as possessing dignity as rational beings.

Kant himself expressly denied that animals and entities of nature had any such moral value and, by extension, that we have any sort of moral duties and obligations towards them. As Kant notes in the Groundwork, animals are "non-rational creatures

[that have] […] only sensuous impulses" (Kant, 1785, p. 50). For Kant, animals are not capable of self-legislating moral laws as humans can because of their total lack of rationality. In addition, as Kant generally states, "if [certain entities] are not rational beings, [they] have only relative value as means, and are therefore called 'things'" (Kant 1785, p. 28). Hence, for Kant, animals are simply "things" that have no inherent moral value and can be treated instrumentally by humans; that is, they are merely means to an end.

Is Rationality Exclusive to Human Beings?

Yet, it would seem that Kantian Deontology's argument for the differentiation and moral exclusion of animals is problematic. The problem stems first from the fact that the Kantian view assumes that humans are inherently and exclusively rational and that no other sort of organism possesses rationality. Indeed, it is on this basis that Kant ultimately distinguishes animals from humans. The immediate and often cited objection brought up against the Kantian moral view of animals along this line is that animals have been known to exhibit rationality in many different sorts of ways. For example, philosopher Tom Reagan argues that a common household dog might be able to reason that barking loudly would prompt their owner to provide them with some form of nourishment and may bark when hungry. In this case and many other similar sorts of instances, animals seem to exhibit all sorts of capacities for reasoning. In fact, Reagan notes, "it would be the height of prejudice merely to assume that man is unique in being able to reason" (Reagan, 1982, p. 750). Consequently, in the Kantian sense, animals should possess some measure of moral worth and dignity in virtue of such capacities.

Nevertheless, as Cohen points out, such objections attributing certain sorts of reasoning to animals fail to recognize the *real* Kantian criterion for moral worth and dignity that distinguishes humans from animals. For Cohen, this is not simply rationality in and of itself, but *moral self-legislation*, that is, the

capacity to understand and act on moral principles. As Cohen puts it, "humans [uniquely] understand that some things, which may be in our interest, *must not be willed*; we lay down moral laws for ourselves, and thus exhibit, as no other animal can exhibit, moral autonomy" (Cohen 1997, p. 772). Such moral autonomy is something that animals cannot possess because notions such as moral "rightness" and "wrongness" are concepts beyond the grasp of animals. For example, Cohen states that common household pets can never grasp that there is a reason for performing or not performing certain actions, other than from the perspective of self-interest. As Cohen states, "my dog knows that there are certain things she must not do—but she knows this only as the outcome of her learning about her interests, [such as by] the pains she may suffer if she does what had been taught forbidden" (Cohen 1997, p. 772). The dog "does not know [and] cannot know that any conduct is wrong;" he cannot know this, that is, simply on the basis that it has some kind of moral dimension to it (Cohen 1997, p. 772). In this sense, then, Cohen argues, concepts of moral rightness and wrongness do not exist in the world of animals. Consequently, animals, unlike humans, do not have the capacity for moral self-legislation required to be intrinsically valuable in the Kantian sense. In the end, for the Kantian, rationality is not the only thing that matters for moral value; it is, rather, the rational capacity to *comprehend* moral principles.

As Cohen himself remarks, this argument for the lack of moral self-legislation in animals seems intuitively true, as amply demonstrated by the way we make moral judgments. For example, most of us would condemn a person who murders others in cold blood, as a morally evil person. Yet, few would go so far as to describe a cow that kills other beings as morally evil, mainly because the cow would not be conceived of as having the same sort of moral consciousness as the murderer. The murderer, most would say, can comprehend that he should not kill his victims. On the other hand, because the cow is not capable of perceiving that what it has done is morally wrong, it is thus not morally accountable for its actions.

Cohen's response in defense of the Kantian view of animals is a good argument as far as it clarifies exactly why rationality *in and of itself* does not qualify animals for intrinsic moral value. Nevertheless, the problem with the Kantian view persists because it seems to hold only if one assumes that moral self-legislation and even rationality, in general, are characteristics intrinsic and exclusive to human beings; that is, traits that have always been essential to human nature and human nature alone. In short, the Kantian view seems to assume that rationality and the capacity for moral self-legislation are totally out of the reach of animals. In the end, Kantian definitions of all animals as non-rational and all humans as rational are assumed to be totally solid and unchangeable.

Such a view of humans and animals seems only to hold according to a biological view of the world, known as the *Great Chain of Being*. The Great Chain of Being was a view of all living organisms that was prevalent before the rise of evolutionary theory. As anthropologist Jonathan Marks states, according to the Great Chain of Being, all organisms, since the dawn of time, have existed in the separate categories of species that we see today (Marks 1995, pp. 5–6). For example, by this view, human beings always existed as a separate and distinct biological species and have not evolved or changed since their inception. Similarly, chimpanzees also would be considered, by the theory of the Great Chain of Being, as a completely separate species that has remained constant since its beginning. By this account, all living organisms observed in the natural world exist in categories that have never changed.

The Kantian view of animals seems to rely on the Great Chain of Being precisely because it places humans and animals into solidly set categories and is literally built on the assumption that humans and animals have never and can never change or evolve. On this view, human beings have always possessed rationality. In addition, on the Kantian view, animals can never evolve into rational beings; they are simply intrinsically bound to be irrational. On this view, animals have no moral status. Ultimately, then, the Kantian view of animals holds only if the Great Chain

of Being is true in the claim that different species of organisms do *not* evolve and simply stay the same over time.

Yet, modern biological theories of evolution effectively disproved the Great Chain of Being as a correct scientific view of the world. According to evolutionary theory, all living organisms are descended from a common ancestor. On the Darwinian concept of natural selection, all the differences in characteristics seen between species today are viewed as having *developed over time.* According to the concept of natural selection, characteristics useful for the survival of a species would be more likely to be passed genetically to future generations of offspring. The consequence of this is that differences in characteristics between species are entirely a result of natural selection. In other words, such differences in characteristics exist simply because they were useful to a species' survival and, thus, were passed down genetically. For example, on the evolutionary view, rationality and the capacity for moral self-legislation are not traits intrinsic and fundamental to human beings alone. Rather, these traits exist among us today solely because they were useful in helping our ancestors survive. In this way, then, evolutionary science displaced the Great Chain of Being theory as an explanation of biological organisms.

The consequences of this for Kantian theory are relatively straightforward. Evolutionary theory implies that characteristics and traits such as rationality and the capacity for moral self-legislation are not intrinsic and exclusive to human beings in particular. Rather, such characteristics are simply the result of processes of natural selection. In fact, such processes might also allow other species of living things to gain the capacity for rationality. For example, it would be quite conceivable under evolutionary theory for cows to have biologically evolved in such a way that they would acquire the cognitive capacities needed for rationality and moral self-legislation. Thus, according to evolutionary theory, it is possible for animals to develop the capacity to reason and comprehend moral concepts. Consequently, the Kantian argument that animals have no moral importance because they cannot reason and morally self-legislate is wrong because such capacities

are not traits that are exclusive to human beings alone.

At this point, one may object that this argument from evolutionary theory poses a problem for the Kantian view of animals. One may acknowledge that evolutionary theory shows that rationality and the capacity for moral self-legislation are not necessarily exclusively limited to humans alone and, theoretically, may be attainable by different species of animals upon evolution. Nonetheless, even with evolutionary theory, one may argue that the Kantian view still applies to animals that *do not currently* possess rationality and the capacity for moral self-legislation. For example, one may acknowledge that it is *possible* for cows to gain the capacity morally to self-legislate through natural evolutionary processes. Yet, at the same time, one may maintain that this does not demonstrate that cows *that currently cannot morally self-legislate* have any sort of moral importance. For example, cows scientifically verified to have neither the cognitive capacity conducive to rationality nor the capacity morally to self-legislate, would have no moral importance on the Kantian view. Indeed, all that the argument from evolutionary theory seems to demonstrate is that it is *possible* for non-humans to develop rationality and the capacity for moral self-legislation and that one does not necessarily have to be human to have moral status. The Kantian argument that animals presently have no rational capacity required for moral self-legislation and, thus, have no moral status, however, still seems to hold. Consequently, evolutionary theory seems compatible with the Kantian view of animals.

In response to such an argument, it is worth noting (and conceding) that the argument from evolutionary theory does not necessarily establish that *all* animals have moral importance under the Kantian moral system. It simply demonstrates that the Kantian moral system cannot maintain that animals can never have moral importance. This is because it is conceivable, as the argument from evolutionary theory shows, for animals to evolve into morally self-legislating agents. In short, animals are biologically capable of moral self-legislation. By demonstrating this, the argument further establishes that the Kantian system cannot

set forth a strict moral distinction between humans and animals that morally excludes all animals. In the end, then, although the Kantian may still consistently maintain that some animals *can* be morally excluded under the Kantian system, the fact remains that the Kantian system's solid distinction labeling animals as non-rational and as morally unimportant and the reverse for humans, does not hold.

This does not mean, however, that the argument from evolutionary theory has no practical relevance for real-life animals today. Rather, the consequences of showing that it is *possible* for animals to possess the rationality needed for moral self-legislation is that certain species of animals that *currently exist* can be regarded in a better moral light. The clearest demonstration of this is by an analysis of the following real-life example. Philosopher Martha C. Nussbaum references a description by primatologist Barbara Smuts of Smuts' friendship with a wolf-dog named Safi. Whenever Smuts felt depressed, she describes how:

> [Safi] approaches, looks into my eyes, and presses her forehead against mine. Then, without fail, she lies down beside me, maximizing contact between her body and mine… As soon as I am supine, she rests her chin on my chest, right on top of my heart, and locks her gaze with mine until my mood shifts. (Nussbaum, 2001, p. 2).

This case is simply one of many that, as Nussbaum states, show that "many types of animals are person-like—capable of intelligence and planning, capable of emotion and responsiveness, capable of awareness of another animal's feelings, capable of recognizing one another and members of other species as individuals, capable of joy, humor, and delight" (Nussbaum 2001, p. 2). Such a case reveals that animals today might have some comprehension of moral values, such as the obligations of friendship.

Now, in response to such real-life cases, the Kantian view of animals would hold that animals such as Safi cannot actually grasp the moral value of friendship because, as Kant himself states in the Groundwork, animals feel nothing but "sensuous impulses"

(Kant 1785, p. 50). In fact, a Kantian might add that Safi's act of comforting Smuts is nothing more than a bare instinct rather than a conscious act of friendship. The argument from evolutionary theory discounts such a view that animals such as Safi are necessarily and completely incapable of comprehending moral values. In this sense, then, evolutionary theory even allows for animals today to be regarded, not necessarily as entities that are totally different from humans, but, rather, as beings capable of many of the same characteristics that Kantians define humans alone as possessing. In essence, this allows for animals' human-like behavior, such as in the case of Safi, to not simply be dismissed, by default, as a consequence of bare instincts and "sensuous impulses." Rather, the argument from evolutionary theory allows for the possibility that such behavior might actually be genuine and, consequently also allows for us to see animals such as Safi in a much better light. Hence, on the basis that, animals are biologically capable of rationality and moral self-legislation, the strict Kantian division between humans and animals breaks down.

Problems with Extending Kantian Dignity to All Human Beings

The previous section explored a problem with the Kantian view of animals. The problem arises out of the fact that it is conceivable for rationality to be extended into the animal kingdom. A second, independent problem with the Kantian view of animals emerges from the opposite direction: the fact that not all human beings have the capacity for rationality and moral self-legislation. Indeed, one of the most common objections to the Kantian view is its implication for numerous categories of non-rational human beings. As one may argue, by the Kantian view, many different categories of human beings lose moral importance, along with animals, because they lack the Kantian criterion for moral value: the capacity morally to self-legislate. This would include small children and those with severe mental handicaps because such people would not be able to rationally comprehend moral prin-

ciples. The problem, one may point out, is that the Kantian view would exclude such categories of human beings, along with animals, as having no moral worth. Thus, the Kantian view allows for the absurd conclusion that it would be permissible to treat them instrumentally as things rather than as persons. The Kantian view of animals, then, one may argue, is wrong because it leads to just such conclusions.

Kantian scholar Allen Wood attempts to defend the Kantian view from this argument with a response similar to most Kantian solutions to the problem. As Wood argues, the key to defeating this argument against the Kantian view is the concept of *potential rationality*. As Wood states, "it would show contempt for rational nature to be indifferent to its potentiality in children, and to treat children as mere things or as mere means" (Wood, 1998, p. 11). In other words, Wood argues that, from a Kantian standpoint, children have moral worth and dignity because they have the *potential* for rationality. Children have the capacity and potential to grow into adults capable of comprehending moral truths. Thus, treating children with such potential as if they were non-rational beings would morally degrade them. One may argue, then, that children actually possess moral dignity and value because they have the potential for rationality and moral self-legislation. At the same time, this is compatible with Kant's view on animals because, unlike children, animals are incapable of gaining the capacity for rationality and moral self-legislation, regardless of how much they are able to grow and develop over their lifetimes.

Wood also uses a similar argument to extend moral worth and dignity to the mentally disabled. As Wood notes, "similar points might be made about respecting rational nature in people who have temporarily lost it through disease or injury [as] [i]t would show contempt for rational nature not to care about them, and to do nothing to help them recover their rational capacities" (Wood 1998, p. 198). In other words, the mentally disabled have already shown that they possess the capability to be rational and to morally self-legislate and, on this basis, possess the potentiality for rationality and moral self-legislation. Consequently, as Wood

notes, not only are we bound by the Kantian duty to respect them, but are also obligated to aid them in recovering their rationality. Thus, as in the case of children, the mentally disabled, from a Kantian view, also possess moral worth and dignity.

Yet, the argument from potential rationality does not seem adequately to answer the objection. The problem begins with the fact that such reasoning only extends moral worth to people who have possessed or will possess rationality at some point in their lifetime. For example, children are conceived as morally valuable because they *will* possess rationality in the future. Likewise, elders with crippling mental diseases, such as Alzheimer's syndrome, are conceived as morally valuable because they *possessed* rationality at some point in their lifetimes before succumbing to the disease. Although such reasoning may cover many categories of non-rational human beings, it does not extend moral worth to mentally disabled human beings who have been cognitively impaired since birth. Indeed, those who are mentally disabled from birth *have never and will never* possess rationality at any point in their lives. For example, those affected by forms of mental retardation have, since birth, little cognitive capacity. The problem, then, is that, even with Wood's response, many mentally disabled humans still are labeled by the Kantian view as having no moral worth.

Of course, a proponent of Wood's overall argument may respond to the problem by arguing that, even in cases in which people have been mentally disabled from birth, there still exists some sort of potential for rationality and moral self-legislation. One may argue that the causes of many forms of mental retardation are biological abnormalities. For example, Fetal Alcohol Syndrome is known to cause mental retardation; children are often born with cognitive impairments and suffer from mental retardation due to the consumption of alcohol by pregnant women. Genetic diseases are another such cause of mental retardation. One may argue that, without such biological abnormalities, those suffering from mental retardation would be normal human beings and like other normal human beings, they would be rational, morally self-legislating agents. Furthermore, the state of existing technology might

eventually make it possible for surgical operations to fix such biological abnormalities and restore cognitive functioning to such beings. Consequently, one may argue that those suffering from mental retardation actually do possess a potentiality for rationality because they can conceivably be cured of whatever causes them to be cognitively impaired. By extension, then, this would mean that they do possess moral worth and dignity.

However, this line of reasoning actually exposes the Kantian position on the issue to even larger and more fundamental problems. Such problems arise with regard to what meaning the concept of potential rationality takes. The best illustration of this is by a thought-experiment. Suppose that scientific researchers managed to find a way to implant a gene into chimpanzees that caused them to achieve the same cognitive state as human beings. Chimpanzees, then, by this procedure, would be able to think and reason in the same way that humans do. They would even be able to comprehend and legislate abstract moral rules and principles. According to the concept of potential rationality, then, would this not mean that chimpanzees would suddenly gain moral worth because of this potentiality for them to attain rationality?

The fascinating truth is that such an odd and outlandish thought-experiment is not so far-fetched from today's world. Today, scientists have already concluded that there is, in fact, an extremely minute genetic difference between chimpanzees and human beings. In fact, for that matter, science tells us that there is, comparatively speaking, very little genetic difference between most animals and human beings. In addition, the field of molecular genetics has made great strides forward in terms of analyzing the content and function of small particles of genetic material. Given such huge scientific advances in terms of our knowledge of genes and molecular genetics, it does not seem so crazy to hypothesize that it might one day be possible to alter genetic makeup to transform non-rational animals into rational, morally self-legislating agents.

The implications of such a scenario are disastrous for Kantian theory. By this scenario, most, if not all, animals in the world

would have moral worth and dignity because it would somehow be *possible* to alter their genetic makeup to transform them into rational, morally self-legislating agents. In fact, it would be just as *possible* as trying surgically to restore the cognitive abilities of the mentally retarded. Consequently, all animals would be entitled to the same rights as the mentally retarded. The problem, then, for Wood's defense of the Kantian view of animals is that, with the advance of science and technology, animals would have the *potential* to become rational agents.

Because of this, the Kantian moral distinction between humans and animals once again breaks down. The problem here is that the concept of potential rationality is crucial for Kantians effectively to lay ground for the conclusion that children and the mentally disabled can be counted as morally valuable beings. Yet, this criterion for moral value, in the end, is ineffective because, when technology becomes sufficiently advanced, animals can also be said to have some kind of potential for rationality. Thus, the Kantian must either concede that *all* humans cannot be considered morally valuable beings (something contrary to common moral intuitions), or be forced to resort to the concept of potential rationality and concede that *both* humans and animals have moral worth. Because of this, then, the Kantian attempt to maintain an effective moral distinction between humans and animals again collapses.

Concluding Remarks

The two areas of examination here highlight the problems of the Kantian moral exclusion of animals, from two different angles. As shown, the problem for the Kantian view becomes first that it is conceivable to extend moral worth into the animal kingdom because, in the end, the capacities for rationality and moral self-legislation are not traits exclusive and intrinsic to human beings alone. Rather, they are traits born through natural processes; processes through which species of animals might similarly develop the capacity for rationality and moral self-legislation.

Consequently, the fact that it would even be *possible* to extend moral dignity to different species of animals completely demolishes the solidity of the Kantian moral distinction between humans and animals.

In addition, a second, almost reverse problem also holds true of the Kantian position: that the Kantian moral distinction actually does not include categories of human beings. Such categories include children and the mentally deficient. As shown, any attempt to resolve this problem using the concept of potential rationality is not sufficient for upholding the Kantian view. Indeed, the concept of potential rationality is simply not an effective means of distinguishing between humans and animals because animals too can be biologically capable of becoming rational agents, such as by genetic manipulation. In the end, then, for this second reason, it is impossible for the Kantian to effectively morally distinguish and exclude animals without also excluding categories of human beings. Consequently, the Kantian moral distinction between humans and animals, again, crumbles.

Such objections, besides exposing flaws in the Kantian view of animals, may also seem to hint at a much more general problem with any moral system that attempts morally to distinguish all humans from all animals. The fact, as evidenced by the analysis of the two problems of the Kantian view discussed here, is that all living organisms share biological similarities and possess the same biological potential to develop and attain a variety of traits. Such biological similarities make it extremely difficult, if not impossible, to maintain a solid moral distinction between humans and animals that can somehow allocate moral dignity to *all* human beings and, at the same time, morally exclude *all* animals. In other words, it becomes difficult to narrow down a particular quality exclusive to human beings alone that justifies the moral exclusion of animals and establishes the moral dignity of all human beings. The Kantian view of animals seems to exemplify this difficulty.

Of course, it is important to note that this, by itself, does not necessarily *prove* that all theories arguing for the moral exclusion of animals are inevitably wrong. All that has been shown here is

that the Kantian view of animals is problematic and that similar theories regarding the moral status of animals may face like difficulties. Ultimately, the arguments here simply pose a challenge for theories advocating the moral exclusion of animals. It is for subsequent theories that advocate the moral exclusion of animals to attempt to solve such difficulties and establish a more stable basis for denying moral worth to animals.

Bibliography

Cohen, Carl. (1997) "Do Animals Have Rights?" Ed. J. Boss. *Analyzing Moral Issues*, pp. 768–775 (New York: McGraw-Hill)

Kant, Immanuel. (1785) *Groundwork for the Metaphysics of Morals*. Ed. Jonathan Bennett. *Early Modern Texts* (September 2008). p. 5–32. 8 February 2010. <http://www.earlymoderntexts.com/pdf/kantgw.pdf>.

Marks, Jonathan. (1995) *Human Biodiversity* (New Brunswick: Transaction Publishers)

Nussbaum, Martha C. (2001) "Animal Rights: The Need for a Theoretical Basis," *Harvard Law Review*, 114, pp. 1506–1549

Reagan, Tom. (1982) "The Moral Basis of Vegetarianism," Ed. J. Boss *Analyzing Moral Issues*, pp. 748–753 (New York: McGraw-Hill)

Wood, Allen W. (1998) "Kant on Duties Regarding Nonrational Nature," *Proceedings of the Aristotelian Society* 72, pp. 189–228

Contributors

Michael Anderson. B.A. Philosophy, Biola University. Metaphysics, Philosophy of Language, Epistemology. Michael intends to apply to PhD programs in philosophy in the Fall of 2010.

Kyle Arneson. B.A. Philosophy, University of California, Los Angeles. Metaphysics, Philosophy of Mind, Philosophy of Science, Neurophilosophy.

Richard Burton. B.A. The Johnston Center for Integrative Studies, the University of Redlands, Redlands, CA. M.A. Philosophy, California State University, Los Angeles. Contemporary Critical Theory, 19th and 20thCentury Continental Philosophy, Post-colonial/Anti-colonial Theory, Violence, Post-Structuralism, Literary Criticism, History and Theory of the Novel. Richard intends to apply to PhD programs in Cultural Studies or Interdisciplinary Studies in the Fall of 2010.

Jasmine Hall. B.A Philosophy, California State University, Los Angeles. Feminist Philosophy, Epistemology, Existentialism, Ethics. Jasmine intends to pursue an M.A in philosophy, then attend law school.

Italia Iossif. B.A. Philosophy, California State University, San Bernardino. Ethics, Meta-ethics, Applied Ethics, Social-Political Philosophy, Philosophy of Law. Italia intends to apply to either Law Schools or PhD programs in philosophy.

Matthew Leonard. B.A. Philosophy, B.A.T. Theology, California Baptist University. Philosophical Logic, Metaphysics, Philosophy of Language, Philosophy of Mind. Matt will attend UC Davis in the Fall of 2010 for his PhD in philosophy.

Carlos Maguina. B.A. Philosophy, UCLA. Philosophy of Language, Epistemology, Metaphysics, and History of Analytic Philosophy. Carlos intends to apply to PhD programs in philosophy in the Fall of 2010.

Ali Mohsen. B.A. Philosophy, California State University, Northridge. Classical Islamic Philosophy, Phenomenology, Metaphysics, Ontology, Philosophy of Mind. Ali is finishing his master's degree and intends to apply to PhD programs in philosophy in the Fall of 2010.

Anthony Ristow. B.A. Film Studies, University Michigan, Ann Arbor. 20th Century Continental Philosophy, Existentialism and Post-Structur-

alism. Anthony intends to apply to PhD programs in philosophy in the Fall of 2010. Anthony is also an aspiring screen-writer.

Raj Shah. B.A. Philosophy, California State University, Los Angeles. Social and Political Philosophy, Applied Ethics, and Metaphysics. Raj will be applying to Law Schools in the Fall of 2010.

April Van Bibber. B.A. Philosophy, California State University, San Bernardino. Social-Political Philosophy, Ethics, Philosophy of Law. April is currently completing her master's thesis at California State University, Los Angeles, and plans to write professionally after graduation.

James Winburn. B.A. Philosophy, San Francisco University. Phenomenology, Post-Structuralism, Aesthetics, Semiotics. James is currently a master's student at California State University, Los Angeles, and plans to apply to PhD programs in philosophy and art theory.

Master of the 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 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