Stance \stan(t)s\: a rationalized position or mental attitude

Featured Authors

Jenna Tomasello
Max Siegel
Kyle Curran
Chad Marxen
Jenny Carmichael
Mark McGinn
Kristen Wells
Vesak Chi
John C. Hill
Oda Storbråten Davanger
and a special interview with
Marilyn Frye

Stance is a recipient of the American Philosophical Association/Philosophy Documentation Center prize for Excellence and Innovation in Philosophy Programs.

Stance: An International Undergraduate Philosophy Journal ■ Volume 6 ■ April 2013

STANCE

An International Undergraduate Philosophy Journal

Stance

An International Undergraduate Philosophy Journal

Stance: An International Undergraduate Philosophy Journal is produced and edited entirely by undergraduate students. We aim to enrich student learning by providing an opportunity for undergraduate students to have their original scholarly work reviewed by, or published in, a peer-reviewed academic journal.

Stance is published annually in April. The deadline for submissions is mid-December. All papers are carefully considered in multiple anonymous reviews. Notification of initial decision is in February. All authors receive constructive feedback concerning submissions.

Having an international standard serial number [ISSN 1943-1880-print; ISSN 1943-1899-electronic], *Stance* is registered with the U.S. Library of Congress. *Stance* is listed in *The Philosopher's Index.*

Further information regarding *Stance* is available at: http://stance.iweb.bsu.edu. Back issues can be found on the Ball State University Virtual Press at: http://www.bsu.edu/library/virtualpress/stance.

Inquiries should be directed to David W. Concepción, Ph.D., dwconcepcion@bsu.edu.

Stance gratefully acknowledges Ball State University's support of the publication of this journal.

Staff

Contents

Sexual Harassment and Objectivity:

Why We Need Not Ask Women If They Are Victims

Jenna Tomasello

Abstract: Sexual harassment is often understood as a subjective notion that asks the woman if she has been victimized. This paper argues that we need not ask women if they are victims by conceptualizing sexual harassment as an objective notion that holds the perpetrator accountable for his actions. In making my case, I will apply an objective conception of sexual harassment to the U.S. Supreme Court case *Meritor Savings Bank v. Vinson* by drawing on the feminist view of sexual harassment given by Anita Superson and the role of equality and autonomy as motivated by Ronald Dworkin and James Griffin, respectively.

Subjective vs. Objective Distinction

Sexual harassment is often identified as a problem only after a woman comes forward to reveal that she has been victimized.[1] I find this conception problematic because it relies on an overly *subjective*[2] interpretation of events rather than *objective*[3] justification. It seems reasonable to suppose that sexual harassment can make women feel uncomfortable or threatened and thus it seems reasonable to suppose that a woman will know when she has been sexually harassed. For this reason, sexual harassment has come to be understood by many as something that can only be subjectively determined. However, I will argue that sexual harassment can and should be objectively determined. In making my case, I will draw on the feminist view of sexual harassment given by Anita Superson and the role of equality and autonomy as motivated by the views of Ronald Dworkin and James Griffin, respectively.

In her paper titled "A Feminist Definition of Sexual Harassment," Anita Superson provides an objective definition of sexual harassment that is counter to the common subjective view. Rather than defining it as annoying, disturbing, or unwanted gestures, actions, threats, or demands, Superson defines sexual harassment as "a form of sexism . . . about the domination of the group of men over the group of women."[4] The subjective definition "puts the burden on the

[1] This paper focuses on sexual harassment that is done to women by men not only because it is by far the most common form, but because this form is fundamentally different from women on men, men on men, and women on women harassment. This difference will be explained in more detail towards the end of the paper.

[2] Subjectivity refers to an individual's perspective, feeling, or belief on a given matter.

[3] Objectivity refers to the reality or truth about a given matter independent of an individual's perspective.

[4] Anita Superson, "A Feminist Definition of Sexual Harassment," *Journal of Social Philosophy*, 24, no. 2 (1993): 400.

victim" to establish whether she was sexually harassed, whereas the objective definition asks "whether the behavior . . . expresses and perpetuates the attitude that the victim and members of her sex are inferior."[5] In other words, the subjective version asks the victim whether she has been harassed, while the objective version suggests that a woman could be harassed and not know it. Using moral concepts like equality and autonomy, this paper argues for an objective conception of sexual harassment in order to hold the perpetrator accountable for his actions, rather than place the burden of proof on the victim.

To illustrate the difference between the subjective and objective conception of sexual harassment, consider the case of *Meritor Savings Bank v. Vinson*. Mechelle Vinson filed suit against her former employer, claiming that over the course of the four years she worked there, she was repeatedly harassed by her supervisor Sidney Taylor. Vinson brought forth five accusations against Taylor. The first incident occurred shortly after she was hired when Taylor invited Vinson out to dinner and while out, suggested they go to a motel to have sex. Vinson initially refused, but eventually agreed for fear of losing her job. After the motel incident, Taylor made repeated demands for sexual favors both during and after business hours in which Vinson estimated that she had intercourse with him some 40 or 50 times. Later, Taylor fondled Vinson in front of other employees. Taylor also followed Vinson into the women's restroom and exposed himself to her. And finally, while alone in the women's bathroom, Taylor forcibly raped Vinson on several occasions. Taylor denied all the allegations, claiming that he never made suggestive remarks, fondled, had sexual intercourse, or even asked Vinson to do so. These accusations that were brought forth ultimately influenced the trial court's subjective reasoning in determining the ruling of this case.

The trial court found that Vinson was not a victim of sexual harassment because her sexual "relationship" with Taylor was "voluntary" based on her "willingness" to participate in the sexual acts. The trial court seems to be appealing to the myth that women "welcome, ask for, or deserve" harassment.[6] On this view, men are justified in catcalling, fondling, or making suggestive sexual remarks because women who flirt or dress "provocatively" should expect such harassing treatment. This myth arises out of a flawed subjective conception of sexual harassment because it questions the behavior of the victim rather than the actions of the perpetrator. Presumably, a woman who playfully tosses her hair or wears a short, formfitting dress cannot find sexual remarks annoying or be disturbed by grabbing because she has invited men to act in that way. Since the subjective version asks the victim whether the treatment was welcomed, sexual harassment becomes a question of whether a woman has good reason to feel victimized.

Vinson appealed the decision and her case eventually made it to the United States Supreme Court where the ruling was reversed. The Supreme Court argued that the trial court "erroneously focused on the 'voluntariness' of [Vinson's] participation in the claimed sexual episodes" and should have questioned whether or not the "alleged sexual advancements

[5] Ibid., 403-5.
[6] Ibid., 402.

were unwelcomed."[7] Contrary to the trial court's tacit endorsement of the "women ask for it" myth, the Supreme Court ruled that Vinson was a victim of sexual harassment because of the "hostile work environment" created by the "unwelcomed" sexual advances made by Taylor.[8] Notice here that even though the Supreme Court recognized that the trial court made a mistake by claiming Vinson was a willing participant, the Court still relied on a subjective conception of sexual harassment to arrive at their verdict.

It is important to note that women often know when they have been sexually harassed. However, I am arguing for objective justification of sexual harassment, rather than subjective, because I think there are cases in which women are mistaken. I believe there are cases of women 1) having been sexually harassed and failing to recognize it, 2) having been sexually harassed and recognizing it, but failing to react properly, and 3) perceiving something as sexual harassment when it was not. For example, suppose, for the sake of argument, there are six coworkers (three male and three female) having lunch together in the employee break room. Their boss enters, interjecting into the conversation by telling a sexist joke that is directed at the three women in the room. The first woman laughs along with her boss and the other men in the room, unaware of the problematic nature of the joke. The second woman nervously laughs along and even though she is angry and offended by the joke, decides not to speak up for fear of making the situation even more uncomfortable. The third woman does not laugh because she, like the second woman, is angry and offended by the joke, but does decide to speak up and express why she finds the joke problematic rather than funny. The three different reactions of the women speak to the problem of identifying sexual harassment by evaluating a victim's reaction. The thought experiment is intended to illustrate that all three women were sexually harassed even though one did not realize, one realized but chose not to respond, and one realized and responded accordingly. It could be inferred by the reactions of the first and second women that sexual harassment did not take place because the sexist joke did not elicit a response indicating that the comment was unwanted. It seems that there are situations in which a woman's reaction is a poor indicator of sexual harassment, and it is because of those situations that I think a victim's reaction is not as important or as good of an indicator as the perpetrator's action.

In sum, *Meritor Savings Bank v. Vinson* is exactly the kind of case that is troublesome because it embodies the problematic nature of the subjective definition of sexual harassment. The trial court held that Vinson *was not* a victim of sexual harassment because of the "voluntariness" of her participation in the repeated sexual incidents. However, the Supreme Court held that Vinson *was* a victim of sexual harassment because the sexual advances, though they were not involuntary, were "unwanted." Both courts approached the issue in subjective terms that asked Vinson whether she was a victim of sexual harassment, and neither asked whether Taylor's actions expressed a view that women are inferior to men. Placing the burden of proof on the victim to prove that she has been sexually harassed perpetuates the existing power structure

[7] *Meritor Savings Bank v. Vinson,* 477 U.S. 57 (1986): 67-8.
[8] Ibid., 63-9.

of the sexes and does not explore the possibility that the perpetrator's actions project the false view that the sexes are unequal in worth. Although I concur with the Supreme Court's ruling, I disagree with their reasoning because I believe that an objective conception should have been applied in determining that Vinson was a victim of sexual harassment. Taylor should have been held accountable for his actions, and his accountability should not have relied on Vinson proving that she was a victim to the Court.

The Role of Equality

In order to illustrate how objectivity can and should be applied to sexual harassment cases, a moral component should be introduced. Ronald Dworkin makes a distinction between rules and principles.[9] Rules are "all or nothing." For example, the legal drinking age is 21, or it is not. It could be the case that the legal drinking age is 18, but it is not because the 'rules' in place state otherwise. Principles, by contrast, lack the precision of a rule. For example, fairness is a fundamental principle that is recognized as valuable, but there is room to disagree about how it applies to a particular case. Rules and principles are both objective concepts; however, rules are sharply defined and principles make up a broad constellation of values.

Consider the role equality plays in the *Vinson* case. According to Dworkin, since equality is a fundamental principle that is recognized as valuable, anything that does not adhere to the principle of equality by suggesting, promoting, or allowing inequality is wrong. It could never be the case that suggesting, promoting, or allowing inequality is right, unless equality is in conflict with some other principle and consequently trumped. Therefore, given any dilemma, if there are two or more principles in conflict, one must weigh the significance of each, decide which is most important, and prioritize that principle over the other(s). Vinson was a victim of sexual harassment because the sexual advances made by Taylor contributed to an inequality among men and women and therefore violated this fundamental moral principle. In Superson's words, suggestive sexual remarks, fondling, sexual intercourse, and rape are ways of expressing a "domination of the group of men over the group of women."[10] These forms of harassment express and perpetuate the idea that women are inferior to men and that the sexes are unequal in worth. Since there are no conflicting principles in this case, and because equality is recognized as good, it is clear that Vinson was a victim of sexual harassment. Vinson was a victim not because Taylor's actions were "unwanted"—it is likely they were—rather, Vinson was a victim of sexual harassment because Taylor's actions violated the principle of equality that is objectively recognized as fundamentally valuable.

Dworkin's objective principle demonstrates that Vinson was a victim of sexual harassment. Taylor's actions were wrong not because they caused specific harm to Vinson, but because they reinforced a belief that women are

[9] Ronald Dworkin, "Is Law a System of Rules?" in *The Philosophy of Law* (Oxford: Oxford University Press, 1977).
[10] Superson, 400.

unequal to men. In Superson's words, "sexual harassment is designed to 'keep women in their place' as men see it."[11] Therefore, sexual harassment violates the principle of equality because it suggests that women are inferior to men. Unlike the reasoning of both courts that relied on subjective notions like "willingness" and "unwanted," the analysis of Dworkin's principle of equality offers objective justification. Taylor violated the principle of equality not because his actions were annoying or disturbing, but because they expressed the belief that the sexes are unequal in worth. Such sexist attitudes not only perpetuate sexism that harms women collectively as a group, but it undermines the principle of equality that we all recognize as good.

The Role of Autonomy

In order to further illustrate how sexual harassment cases can and should be objectively determined, I will now shift the discussion from moral principles to human rights. James Griffin explains the concept of human rights as being grounded in two ideas: personhood and practicalities.[12] He believes that the ability to make self-governing choices (i.e., autonomy), not have others interfere with our choices (i.e., liberty), and have minimal provisions guaranteed in order to make choices (i.e., food, water, shelter, etc.) is necessary for personhood. Griffin offers practicalities as a way of "drawing the line" between which rights warrant protection and which do not. For example, according to Griffin, liberty is a right necessary for personhood, and rights derived out of liberty include rights of speech, religion, and assembly. However, in the United States, freedom of press is typically included in the above mentioned list of freedoms, but unlike the other rights, it is not a necessary right for personhood in all places of the world. As a matter of practical consideration, in the first world, freedom of press might be a necessary liberty for personhood, but it is likely not necessary in the third world. This is because what it means to be a functioning human in the third world is very different from what it means to be a functioning human in the first world. Similarly, the health needs of men and women are going to look very different given practicalities. It would not be necessary that all people have access to a pap smear or a prostate examination. Given anatomy and the different health needs of men and women, it would not make sense to provide men with pap smears and women with prostate examinations. This is because what it means to be healthy and functioning as a man is different from what it means to be healthy and functioning as a woman.

According to Griffin, autonomy, liberty, and minimal provision are needed for humans to function as humans. Vinson was a victim of sexual harassment, not because Taylor's actions were "unwanted," but because his actions undermined Vinson's personhood status. Taylor violated Vinson's right to autonomy because his actions had an impact on how she was able to govern her life. Even if Vinson was "willing" to participate in the sexual

[11] Ibid., 402.

[12] James Griffin, "First Steps in an Account of Human Rights," in *On Human Rights* (Oxford: Oxford University Press, 2008).

acts with Taylor, her ability to make an autonomous decision was obscured because she was afraid she might lose her job. When a power imbalance exists—as it always does in an employer/employee relationship—there is a level of manipulation available to an employer to abuse. Taylor was able to use his position of power to his advantage because even though Vinson "agreed" to participate in the sexual "relationship," she was not making a truly autonomous choice to do so. Vinson did not freely choose; rather, she was coerced into the sexual relationship because she had to weigh the consequences of the sexual relationship with the consequences of losing her job. As a matter of practical consideration, Taylor's actions did not allow for Vinson to freely exercise her autonomy given the needs of a female employee working under the supervision of a male employer. Furthermore, since Vinson was unable to act as an autonomous agent and a right to autonomy is necessary for personhood, it follows that she did not meet Griffin's minimal requirement for personhood that all people are owed in virtue of being human.

Vinson was a victim of sexual harassment because her personhood was undermined when her right to autonomy was unreasonably restricted. As humans, our autonomy is always restricted to a certain degree to ensure that we do not harm others. For example, I do not have the freedom to murder my ex-lover even though I might want to, because restricting my freedom to kill another person is a reasonable restriction of my autonomy. Murdering my ex-lover is not going to inhibit my ability to function as a human agent. Similarly, employers are able to reasonably restrict the autonomy of their employees by requiring that they show up to work on time, complete their designated tasks, refrain from revealing the business secrets of the company, etc. Notice how not one of these restrictions hinders an employee's ability to function as a human agent. An unreasonable restriction of an employee's autonomy would be requiring the employee to work more hours per week than the maximum allotted, perform dangerous tasks unrelated to the job, or participate in a "hostile work environment." Clearly, any of these restrictions can affect an employee's ability to function as a human agent, whether the worry is death or abuse in verbal, physical, or sexual forms. This case demonstrates an unreasonable restriction of Vinson's autonomy because she was forced to weigh the consequences of a sexual relationship to losing her job at the bank. Vinson understood her participation in the sexual acts could ensure or inhibit her ability to function as a human agent because a job is going to ensure that minimal provisions are attainable.

Does Sex Matter?

It is important to consider cases in which women sexually harass men, men harass men, and women harass women. While I am not trying to suggest that such cases do not exist, sexual harassment in which the woman is the victim and the man is the perpetrator is fundamentally different from these other forms. It is true that regardless of Vinson's sex, the principles of equality and autonomy would have still been violated, but the fact that Vinson was a woman and Taylor was a man is important because it illuminates

the troublesome power structure that exists between the sexes. Superson states "women cannot harm or degrade or dominate men as a group, for it is impossible to send the message that one dominates if one does not dominate."[13] While I agree with Superson that women cannot remind men that they are inferior because of their sex since society is not structured in such a way, I do not want to discount other forms of sexual harassment.

Individuals, regardless of their sex and the sex of the harasser, can be sexually harassed. Superson's feminist definition of sexual harassment defines it as the "the domination of the group of men over the group of women."[14] Her definition highlights that sexual harassment of women by men has a profoundly negative impact on women as a group. Not only does sexual harassment harm the individual, it harms all women because it perpetuates the false belief that women are inferior to men. The *Vinson* case highlights how Taylor's actions not only caused specific harm to Vinson by undermining her autonomy, but also expressed an inequality among the sexes that suggests men are superior to women. Other forms of sexual harassment do not have this same impact. Yes, the individual being harassed is harmed, but the group to which the individual belongs is not being harmed because the harassment does not derive from the sexist attitudes associated with one's membership in a particular group. Superson acknowledges that "if the sexist roles predominant in our society were reversed, women could sexually harass men."[15] Therefore, the reason she denies other forms of sexual harassment is because they are not going to have the same impact given the sexist roles that exist in society. Ultimately, what is most important about Superson's definition of sexual harassment is that it is based in the ability for one group to dominate another group, rather than the ability for one individual to harm another individual.

In sum, sex does matter in this case because the type of sexual harassment that is done to women by men is significantly different from other forms due to the impact it has on women collectively as a group. While that is not to say other forms do not exist, it is important to note that these forms do not elicit harm beyond what is faced by the individual being harassed. This is because society is not structured to evoke the domination of the group of women over the group of men or the domination of either sex on themselves.

Conclusion

Sexual harassment is typically understood as a subjective notion that asks the victim whether she has been harassed. More often than not, a woman is aware when she is a victim of sexual harassment, but then the burden rests on her to prove that she has been victimized. In Mechelle Vinson's case, both the trial court and the Supreme Court relied on a subjective notion of sexual harassment to arrive at their verdicts. Unlike the trial court, the Supreme Court was right to rule in favor of Vinson, but an objective conception of

[13] Superson, 403.
[14] Ibid., 400.
[15] Ibid., 403.

sexual harassment should have been applied. Together, Superson's feminist definition of sexual harassment, Dworkin's moral principle of equality, and Griffin's right to autonomy demonstrate how an objective understanding of sexual harassment should have been applied in determining the verdict of the *Vinson* case. Sexual harassment can be determined independent of subjective interpretation, and this is why an objective conception suggests that we need not ask women if they are victims. ❖

Revising the Principle of Alternate Possibilities

Max Siegel

Abstract: This paper examines the position in moral philosophy that Harry Frankfurt calls the Principle of Alternate Possibilities (PAP). The paper first describes the principle as articulated by A.J. Ayer. Subsequently, the paper examines Frankfurt's critique and proposed revision of the principle and argues that Frankfurt's proposal relies on an excessively simplistic account of practical reasoning, which fails to account for the possibility of moral dilemmas. In response, the paper offers a further revision of PAP, which accounts for Frankfurt's critique, moral dilemmas, and the challenge of causal determinism.

A highly contested position within moral philosophy and the free will debate is the Principle of Alternate Possibilities (PAP), which holds that "a person is morally responsible for what he has done only if he could have done otherwise."[1] At first glance, this principle seems intuitive—it accords with our view that coercion exempts one from moral responsibility. However, PAP is actually quite problematic, as it seems that individuals can be morally responsible for certain actions, even when they could not have done otherwise. In this paper, I will discuss the problems of PAP, explain one prominent proposed revision to the principle, and also offer an alternative position that better accords with our considered moral judgments.

This paper proceeds in five sections. I begin by discussing A.J. Ayer's defense of PAP and articulating the apparent merits of his position. Second, I explain Harry Frankfurt's well-known challenge to PAP, which I illustrate through counterexamples. Third, I explain Frankfurt's proposed revision to PAP. Fourth, I criticize Frankfurt's proposed revision, showing that it presents an excessively simplistic picture of practical reason and thus fails to capture our intuitions about responsibility in moral dilemmas. I suggest an alternate revision that might be more successful. Finally, I address several possible objections to my view.

In his "Freedom and Necessity," A.J. Ayer attempts to save moral responsibility from the challenge of determinism by claiming that moral responsibility requires not freedom from causal determination but rather freedom from constraint.[2] First, Ayer denies the relevance of causal determinism to moral responsibility. His argument is as follows: according to those who invoke causal determinism, we are not responsible for our actions if they were determined according to causal laws. At the same time, we are not responsible for our actions if they were not determined according to

[1] Harry Frankfurt, "Alternate Possibilities and Moral Responsibility," *The Journal of Philosophy* 66, no. 23 (1969): 829.

[2] A.J. Ayer, "Freedom and Necessity," in *Free Will*, ed. Gary Watson (Oxford: Oxford University Press, 1982), 22.

causal laws, as such actions must be the result of chance. As such, anyone who attempts to defend moral responsibility by denying the thesis of determinism will fail to defend moral responsibility. As an alternative, then, Ayer tries to tie moral responsibility to a different sort of freedom: freedom from constraint, which implies the existence of alternate possibilities. On this account, an agent B acts freely when the following conditions are met: (1) if B had chosen to act otherwise, she would have done so, (2) B's action was voluntary insofar as her deliberation was efficacious and not constrained by a psychological abnormality (e.g., kleptomania), and (3) nobody compelled B (e.g., through coercion) to act as she did.[3]

To this account, the hard determinist might object that condition (1) is insufficient because B is causally determined to choose as she does and thus never could choose to act otherwise. Putting aside this objection, as it does not address the legitimacy of PAP but rather the existence of alternate possibilities, we can see that Ayer's account is grounded in PAP. The ability to choose and do otherwise, the freedom of one's deliberation, and the absence of constraint all refer to B's freedom to choose to ψ rather than to ϕ, where ψ represents any action aside from ϕ. When B lacks the freedom to ψ, she is not responsible for ϕ-ing. Condition (1) would be violated, as B could not ψ even if she chose to. Condition (2) may be violated, if B cannot ψ because of a psychological abnormality. Condition (3) may also be violated, if B's inability to ψ is a result of another's coercion. We can thus see the appeal of PAP, insofar as conditions (1), (2), and (3) appear to conform to our intuitions regarding moral responsibility. A failure of any condition appears to exempt B from responsibility.

However, Harry Frankfurt poses a robust and effective challenge to PAP, which weakens Ayer's position. While Frankfurt presents several counterexamples, I will focus on the one that I find most effective. In this example, Black wants Jones to perform a particular action. He is willing to use force in order to ensure that Jones performs this action, but Black chooses not to get involved unless necessary. As it turns out, Jones wants to perform the action that Black desires. Without any knowledge of Black's presence, Jones performs this action. In this case, Jones could not have done otherwise. Black was ready to use coercion if Jones deviated from the desired course of action. However, Jones shared Black's desire that this action be performed and thus performed the action independently, without the influence of an external constraint. Here, Ayer's condition (1) is violated. If Jones had chosen to do otherwise, he would not have been able to do so, as Black would have forcefully intervened. We might say that condition (3) was also violated; Jones was subject to another's constraint and potential compulsion even though he did not act as to test the constraint. However, Jones is still morally responsible, even though he could not do otherwise. A constraint was in place, but Jones' action was fully "his own," so to speak. PAP is thus a flawed principle. Even when an agent cannot do otherwise, she can still be responsible for acting as she does.

[3] Ibid., 22.

As PAP is very much embedded in our moral discourse, Frankfurt offers a revised version that aims to correct PAP's flaws. He suggests "a person is not morally responsible for what he has done if he did it only because he could not have done otherwise."[4] His argument for this revision is twofold. First, the revision solves the problem of the original version of PAP by requiring that the lack of alternate possibilities actually affect a person's actions and not merely lurk as an unconsidered background condition. In the above example, Jones would be morally responsible because even though he could not have done otherwise, this fact was unknown to him and thus affected neither his practical reasoning nor the causation of his action. Second, the revision solves the problem presented by an alternate candidate revision—"a person is not morally responsible for what he has done if he did it because he could not have done otherwise," identical without the word "only"—by forcing us into a reasons-based reading of "because," rather than a merely causal reading.[5] When one invokes the absence of alternate possibilities as an excuse, one typically means that one was forced to act against one's settled reasons. As Frankfurt puts it, one who invokes PAP typically means "when he did what he did it was not because that was what he really wanted to do."[6] B is not exempt from responsibility if he wants to ϕ and is coerced into doing so on top of his preexisting desire. When we invoke PAP, it is because we believe that the absence of alternate possibilities was the operative factor leading to B's ϕ-ing, and Frankfurt's principle appears to capture this intuition.

However, Frankfurt's position suffers from a critical flaw. As above, Frankfurt claims "a person is not morally responsible for what he has done if he did it *only* because he could not have done otherwise" (emphasis added).[7] However, this principle fails to account for the possibility of moral dilemmas: cases in which reasonable moral claims compete and cannot both be satisfied. Frankfurt mistakenly assumes that a person can have no reason for acting against her freely chosen course of action, but this view is incorrect. Consider the following case, a classic trolley problem. Martha must choose between pushing a fat man onto the railroad tracks to save five railroad workers from an oncoming train or letting the fat man live while the railroad workers die.[8] Having studied some moral philosophy, she is familiar with both the Kantian position, which claims that the fat man is inviolable and that Martha ought not to push him, and the utilitarian position, which claims that Martha cannot prioritize any individual's happiness and ought to save the five railroad workers to maximize utility. Martha finds both positions compelling and believes that both provide normative reasons. Indeed, during her deliberation, she walks towards and away from the fat man several times, reversing her position, as she is actually motivated by both positions. The Kantian and utilitarian arguments both appear

[4] Frankfurt, 838.
[5] Ibid.
[6] Ibid.
[7] Ibid.
[8] The "fat man" formulation of the trolley problem is due to Judith Jarvis Thomson, "Killing, Letting Die, and the Trolley Problem," *The Monist* 59 (1976): 204-17.

to give her normative and motivating reasons. After lengthy deliberation, she decides not to push the fat man and walks away. As she is leaving, a gunman approaches her and orders her to push the fat man. Martha pushes the fat man because she cannot do otherwise. In this case, the effect of coercion exempts Martha from moral responsibility. Her action went against the balance of her reasons. However, it does not seem that she pushed the fat man *only* because she could not do otherwise. We saw above that she had normative reasons to follow the utilitarian route and that she was even motivated by these reasons. Normative reasons do not drop away when one chooses to act against them. One can maintain a reason to ψ even after she has decided to ϕ. Frankfurt's principle cannot exempt Martha from responsibility because she had two reasons for pushing the fat man: a considered moral judgment and a need to avoid death, which made her unable to do otherwise. Frankfurt's principle thus fails to capture cases of moral dilemmas.

As such, I propose yet another revision of PAP. In this version, we view PAP as a counterfactual, which draws somewhat closer to Ayer's position as I articulated it above. In essence, PAP should absolve one from moral responsibility when the fact that a person could not do otherwise tipped the balance of reasons in favor of an action that the person would not have performed absent the constraint. For example, like Martha, a person might have strong reasons to ϕ but stronger reasons not to ϕ; she thus decides not to ϕ. If coercion then tips the balance in favor of ϕ-ing, she is not morally responsible, even though she had some reasons to ϕ that emerged from her previously free deliberation. The revised principle is thus as follows: a person is not morally responsible for what she has done when (1) she acted as she did because she could not have acted otherwise, (2) her action went against the balance of her normative reasons aside from the constraint that made her unable to act otherwise, and (3) her action was substantively different from what she would have done absent the constraint. The first condition accounts for Frankfurt-style cases, the second ensures a reasons-based reading of "because" but without taking Frankfurt's simplistic view of practical reason, and the third accounts for true irrationality, weakness of the will, and other cases where people act against the available reasons. The second and third conditions together also account for those who wanted to ϕ and whose actions were subsequently overdetermined by coercion to ϕ. This principle requires that one could not do otherwise, that this constraint was a factor in one's practical reasoning, and that this constraint was decisive against the balance of other available reasons. The constraint need not be the only reason in favor of ϕ–ing, but it must tip the balance in favor of ϕ–ing. By revising the principle in this way, we maintain that one is not responsible for coerced acts, correct for the counterexamples presented by Frankfurt, avoid a purely causal reading of "because," and allow for the possibility of moral dilemmas.

Some might object to my position on the grounds that previously rejected reasons cannot retain any motivational force for an agent. In the trolley problem, Martha's having rejected utilitarian reasoning means that she could not have been motivated by such reasoning when she subsequently chose to push the fat man onto the tracks. However, we can modify this case to see that

this objection is misguided. In a modified version of the case, Martha initially favors the utilitarian position over the Kantian position, but she is weak of will and thus fails to push the fat man onto the tracks. She turns to walk away and is confronted by the gunman, who orders her to push the fat man. In this case, Martha might feel a sense of relief; she might be grateful for the approach of the gunman, which enabled her to overcome her *akrasia* and act in accordance with her considered moral judgment. In this case, Martha acts for the dual reasons of coercion and her belief that she has a duty to maximize utility. If we ask her why she acted as she did, she would point to both of these reasons. Frankfurt's revised version of PAP would call Martha morally responsible for her actions in this case, since she acted as she did both because she could not do otherwise and because she wanted to comply with utilitarianism. However, Martha's action was not truly "her own." Absent coercion, she would not have acted as she did. As such, this revised case confirms that one can be motivated by previously rejected reasons, even if coercion is an important factor in one's practical reasoning, and suggests that we should favor my revision of PAP over Frankfurt's.

A further objection to my view might be that Ayer's position accounts for cases like Martha's and, on one reading, is also compatible with Frankfurt's critique. One might say that Ayer's first condition—that if B had chosen to act otherwise, she would have done so—is compatible with Frankfurt's principle if it is read as a sufficient rather than a necessary condition. We see from the Jones case that this condition can be violated, and one can still be morally responsible. However, if this condition is met, we rule out the possibility of exemption by lack of alternate possibilities. One who could act otherwise but does not do so is not even exempted by the original version of PAP. The second condition—that B's deliberation was effective—is unaffected by Frankfurt's principle. When we exempt someone psychologically abnormal from moral responsibility, we are not appealing to their inability to do otherwise but rather to the fact that their deliberation has been impaired all along. Even if psychologically impaired person B could have chosen to ψ rather than to ϕ, B would still not be responsible. The third condition—that B's action was not constrained or compelled by another person—aims at the same effect as Frankfurt's principle and requires only moderate revision. We saw above that in Jones' case, condition (3) was violated, but Jones was still responsible. However, if we revise condition (3) to state that the constraint or compulsion by another was the only reason that B chose to ϕ, we end up with Frankfurt's result. If we instead say that the constraint or compulsion was an operative reason that B chose to ϕ, some might say that we also account for the case of Martha.

Ayer's position, however, is subject to the challenge of causal determinism in a way that my position is not. As above, Ayer's first condition for moral responsibility is that if an agent had chosen to do otherwise, she would have done otherwise. However, if determinism is true, an agent will never be able to choose or do otherwise. This counterfactual, then, will be irrelevant, and Ayer's position will leave psychological abnormality and coercion as

the only two conditions that could exempt one from moral responsibility. However, these are not the only two exempting conditions available to agents. Some agents act as they do and are exempt from moral responsibility because of physical disability or situational constraints, among many other factors. Revising Ayer's position to make it a suitable alternative to PAP would require developing a long list of exempting conditions and result in a clumsy, highly complicated principle. For theoretical simplicity and to avoid the challenge of causal determinism, we should opt for my position.

We have seen that the principle of alternate possibilities has great appeal but is deeply flawed. Though it requires some revision, PAP captures an important intuition about moral reasoning: namely, that constraint and compulsion can absolve one of moral responsibility, albeit only in specific circumstances. The lack of alternate possibilities is not sufficient to exempt one from moral responsibility, but when this constraint substantively impacts one's actions, a revised version of PAP may grant one exemption. When reformulated as I propose, PAP can effectively distinguish between cases of responsibility and cases where constraint voids responsibility. Though critical of both Ayer and Frankfurt, my position recognizes the merits of both and corrects for possibilities that they overlook.[9] ❖

[9] I owe many thanks to Professor Victoria McGeer for her guidance in the preparation of this paper and to anonymous referees for their helpful comments on an earlier draft.

Change and Moral Development in Kant's Ethics

Kyle Curran

Abstract: This paper is concerned with an ambiguous aspect of Kant's ethics, namely, how moral change is possible. Kant conceives that change is possible, indeed desirable, without making clear the mechanism by which this change occurs. I conclude that one's moral development must come about through the autonomous rationality of humanity. This allows for the moral law to be held at all times and for the rejection of immoral sentiments and inclinations. Further, it is constant soul-searching that allows one to keep a check on their maxims, facilitating the development of a moral disposition.

Personal change and moral development is a goal for many individuals.[1] Yet there seems to be controversy surrounding how we achieve this desired end. Even within the works of Immanuel Kant, potential answers are vague and ambiguous. This paper will address this problem within the Kantian ethical system, seeking to offer a reading on how one may morally develop within this program. Certainly change is desirable for Kant, but by which means are individuals to acquire this moral goal? Kant's ethical theory relies upon reason as a metaphysical groundwork by which all moral actions are to be judged, allowing for uniformity and universality in the moral law.[2] If we follow Kant's understanding of morality and evil as well as his premises for human behavior, it would appear that personal change is made possible by appealing to this foundation of reason.[3] I argue that, based on Kant's conception of morality and evil, people have it in their capacity as rational agents to alter their behavior and will to accord with the categorical imperative, or, conversely, to become evil if they fail. Because of people's autonomous rational nature, actions are not only, in the strictest terms, dictated by mechanical bodily desires. At least in theory, Kant allows for individual moral change by appealing to humanity's

[1] Personal change and development in this paper refer to the transition of a rational adult from a state of immorality to one of morality, and vice versa. Specifically, use of 'positive change' refers to the change from immorality to morality, while 'negative change' refers to the change from a moral to immoral status.

[2] Immanuel Kant, *Grounding for the Metaphysics of Morals*, Book I in *Ethical Philosophy: The Complete Texts of Grounding for the Metaphysics of Morals and Metaphysical Principles of Virtue (Part II of the Metaphysics of Morals) with On a Supposed Right to Lie Because of Philanthropic Concerns,* 2nd ed., trans. James W. Ellington, Book I (Indianapolis/Cambridge: Hackett Publishing Company, 1994), Ak. 401, 413-4.

[3] Because this project concerns a challenge to Kant's ethics, I have taken reason to refer to that internal thinking process by which one can grasp abstract principles *a priori*.

possession of autonomous rationality, or in the case of negative change by admitting bodily desires into our maxim, "the subjective principle of volition."[4]

The controversy is not so much that people can change, but rather how, and what this mechanism by which we change is. Kant states:

> a change of mentality is an exit from evil and an entry into the good, the putting off of the old human being and the putting on of the new one . . . this change as an intellectual determination, however, does not contain two moral acts separated by an intermediate time, but is only a single act, because the abandonment of evil is possible only through the good attitude that brings about the entry into the good, and vice versa.[5]

This passage is explicit: change is possible. In fact, Kant seems to imply that we ought to morally change, and yet he leaves ambiguity regarding which process one must use to achieve this moral development. I intend to interpret this Kantian conception of moral development and demonstrate the means by which this desired change may be achieved—means intimately connected to his conception of autonomous rationality.

By arguing this stance, I aim to contribute a number of ideas to a wider Kantian program: that failure in duties, whether in implementation or recognition, does not imply that one does not have moral capacity; that hope remains, insofar as people are rational agents, to change for the better; that morality, as an intrinsically individual exercise, relies significantly upon that old adage 'know thyself'; that despite one's personality and inclinations, morality is always in our capacity; and finally, that, despite uncertainties one may have over one's maxims, these same uncertainties are a strength in that such reservations add to a constant soul-searching and pondering of the moral laws. Neglecting to ponder one's moral choices results in either an unwarranted arrogance in our moral capacities or a disregarding of the moral laws entirely.

In the end, I believe the idea of change is fundamental to ethics as a whole, for it shapes our perceptions on whether we can be moral at all. Such an idea is foundational to theories of how people ought to act. If people ought to act in certain ways, change must lie at the core, for if people could not change to meet this "ought" then ethical theories would cease to have any true relevance to human behavior. Change as a concept deserves exploration as a crucial aspect of ethics. By explicating what I take to be Kant's mechanism for moral change, I intend to demonstrate how moral change is possible, both in Kant's ethics and in general.

I have already alluded to the double-sided nature of moral change. It is entirely possible that a morally corrupt person may become morally in line with the categorical imperative, "an action . . . objectively necessary in itself,

[4] Kant, *Grounding*, 13, note 13.

[5] Immanuel Kant, *Religion within the Bounds of Bare Reason*, trans. Werner S. Pluhar (Indianapolis/Cambridge: Hackett Publishing Company, 2009), Ak. 74.

without reference to another end," just as a morally righteous person may become corrupted and 'evil' in a Kantian understanding.[7] Both are in line with Kant's program. It is worth noting that the type of moral change I refer to is not directed toward one end alone. Further, it is pertinent to note that reaching a perfect Kantian ideal does not make one impervious to corruption.

In order to establish this capacity to change, I refer to Kant's *Grounding for the Metaphysics of Morals* and "Duties to Oneself" to demonstrate Kant's description of an individual moral agent's rational autonomy. Kant argues in the *Grounding* that "reason must regard itself as the author of its principles independent of foreign influences . . . this is to say that the will of a rational being can be a will of its own only under the idea of freedom, and that such a will must therefore . . . be attributed to all rational beings."[6] Furthermore, although people have a bodily nature, it is through rationality that people are free, "for independence from the determining causes of the world of sense (an independence which reason must always attribute to itself) is freedom. The idea of freedom is inseparably connected with the concept of autonomy, and this in turn with the universal principle of morality, which ideally is the ground of all actions of rational beings."[7] The result is that through rationality people are able to free themselves from the senses. Without rationality, people would not be free, only acting mechanically and as animals. Kant notes that "all animals have the faculty of using their powers according to will. But this will is not free. It is necessitated through the incitement of stimuli, and the actions of animals involve a *bruta necessitas*."[8] Kant further notes "man alone is free; his actions are not regulated by any subjectively necessitating principle."[9] The restriction to this freedom is the universal law and the ignoring of one's inclinations so that one may live in line with the "essential end of humanity."[10] Therefore, people have it in their capacity, according to Kant, to isolate this rational aspect and act according to its precepts. This is crucial for our understanding of how people can change within a Kantian moral system.

Perhaps this is still too metaphysical an explanation to illustrate how this change is possible, and a more specific look at what Kant calls duties is needed. Kant notes that there are both perfect duties (specific actions) and imperfect duties (general behavior) which people ought to strive for based on the categorical imperative.[11] Let us consider perfect duties by examining the act of charity. As the focus is change, I will begin with positive change and suppose there is a person who is not charitable trying to become so. Kant considers the unsympathetic person who, it is presumed, would not donate to charity based on inclinations, but through the appeal to rational autonomy, can do so insofar as he remains a rational agent.[12] Supposedly, through reason

[6] Kant, *Grounding*, Ak. 448.
[7] Ibid., Ak. 452-3.
[8] Immanuel Kant, "Duties to Oneself," in *Lectures on Ethics*, trans. Louis Infield (Indianapolis/Cambridge: Hackett Publishing Company, 1963), 121.
[9] Ibid.
[10] Ibid., 122.
[11] Kant, *Grounding*, Ak. 422-4.
[12] Ibid., Ak. 398-9.

people are able to grasp principles of motivation without necessary reliance upon sentiments of any kind, indicating potential for moral change and the potential to avoid these would-be immoral influences.[13]

However, Kant's conception of human nature, and by extension motivation, seems biased if we consider the question of differing psychologies. It is not clear that rational capacity corresponds to action. For example, consider those with social anxiety. Kant seems to suggest that it is quite simple to conceive of an individual who rationally grasps duties and then moves to implement these by appealing to dictates of reason.[14] However, it is unclear how Kant understands those with anxieties within social situations, where the duty is to act in a way that would put said individual in a stressful situation where he is unable to complete the duty altogether. There would appear to be a disconnect between rationally grasping the moral laws, which Kant may be correct about, and putting these into practice. Simply, this can be formulated as a person who rationally deduces his duties but is unable to complete them despite understanding the implications. In what may be described as a physical entrapment, the will would not be able to express itself through duties despite having what in this case may be a perfect rational understanding of the duties. This indicates a possible problem with Kant's program.

This is no easy comparison to our aforementioned unsympathetic character, for unlike a person devoid of emotion, the person with anxiety can deduce by reason the moral law, but cannot put it into practice. Despite this, Kant's program is not entirely lost. Retaining rational capacity and access to the categorical imperative, these individuals may well strive towards the Kantian imperfect duty to bring one's self and behavior closer
to morality. Although potentially failing in perfect duties, these persons, by still having the capacity to reason, may still aim to perfect themselves in line with the moral law. Although anxious, they may still strive to act in cases where they are uncomfortable. While they may still fail, in many ways they are not alone. I find it difficult to believe that even the most rational and able person would be able to live up to his or her duties at every instance. Thus, these cases remain as nothing more than what seems to be the norm: people with weaknesses looking to make themselves better for the world. Therefore, they still fit into a broad understanding of human nature. What at first might have appeared to be a failing is ultimately compatible with Kantian morality insofar as the subjects have reason and are able to work at perfection and morality. While potential failure is present in certain situations, people are able to strive for change and moral development as long as they are rational agents. Therefore, change is very much possible by using reason as a benchmark for long-term behavioral developments.

In both cases, change is possible through rational autonomy. This is predicated on Kant's ideas regarding reason and its impervious nature to

[13] I have used the term 'sentiment' to refer to that individual feeling of how one wants to act, emotional desires, feelings or other principles of self-love in one's volition. Sentiment is therefore opposed to the categorical imperative and is non-objective. What one wants to do is irrelevant when following the categorical imperative, which is the sole indicator of morality.
[14] Ibid., Ak. 453-4.

disregard immoral inclinations completely.[15] On this point I agree with Kant. I believe it is true that people have a rational aspect, and it is this rational aspect that prevents people from acting as animals, i.e., only by reference to stimuli. This point can be illustrated by referring to the phenomenon of hunger strikes. My intention here is not to argue the ethics surrounding such protests, but only to use these actions as an illustration of this point. If humans were merely compelled by stimuli and bodily needs, hunger strikes would not exist as a human phenomenon, because people would never choose to deliberately starve themselves despite access to food. This phenomenon is only possible if humans have a rational capacity to operate outside of bodily demands and do things from a strictly rational maxim. This point helps illustrate that, since the rational capacity remains autonomous from sentiment and it is this rationality upon which universal morality is based, people can appeal to reason to change for moral development.

I have largely considered moral change in my foregoing analysis, but now I must refer to immoral change, that is to say, from a state of morality to one of depravity. Kant, in *Religion within the Bounds of Bare Reason*, defines evil as a person who is "conscious of the moral law and yet has admitted the (occasional) deviation from it into his maxim."[16] Further, Kant notes that this capacity for evil is natural to humans insofar as all humans have it in their capacity, as embodied beings, to include sentimental inclinations in their maxims.[17] Kant understands that "all evil in the world springs from freedom . . . free beings can only act regularly, if they restrict their freedom by rules,"[18] and that it is the "inner baseness, and not the consequences" which is the principle of evil.[19] On Kant's account, evil arises because of the admittance of non-objective volitions into one's maxims.[20] In essence, this action is to ignore the objective law which reason guides us to. To refer back to the unsympathetic person, let us suppose that she has been consistently charitable in reference to the moral law as incentive. If this person should falter in her morality and not give to charity, instead using her money for a luxurious dinner or for some other personal reason, this would be an evil. Like morality, evil results from our maxims or reasons for doing something, not necessarily the consequences of action. Where morality arises from adopting objective incentives, evil arises from non-objective sentiments.

Thus, the question of whether or not people can become evil is answered with a resounding yes: immoral change can be accomplished via inherent human subjectivity or non-objectivity. Kant admits of no doctrine of 'once saved, always saved' and, as is noted in the *Religion,* deviation from the moral law is "necessary in every human being, even in the best."[21] It is clear

[15] Ibid., Ak. 448.
[16] Kant, *Religion*, Ak. 32.
[17] Ibid.
[18] Ibid.
[19] Ibid., 125.
[20] 'Non-objective' refers to nothing more than the rejection of the universal Kantian law and categorical imperative. This is opposed to the objective and universal Kantian moral law.
[21] Kant, *Religion*, Ak. 32.

that change occurs in a negative sense. All humans, being embodied, have non-objective incentives which they can succumb to. As long as a moral person remains a physical person, he is subject to inclinations that he must always resist through rationality, implying that there is potential for evil if he should fail.

It is at this point I must make a note of the type of change this implies. Both changes I have elucidated are internal changes, which are changes only in our maxims. If a person gives to charity because she wishes to be seen as charitable or wants to impress her peers, this would be a sentimental reason. To be truly moral, she must act objectively. Therefore, change may only be accurately judged by the individual. Since people have rational autonomy, they are able to reflect on their motives and change to become internally moral if need be. People can change insofar as they are individually able to change their internal thinking process. Consider if the opposite were true, and people were conscious of their thoughts but could not change them. The implication would be that the entire Kantian program is false. Also, it would imply that human actions and maxims were completely determined by forces outside of a rational agent's control. However, I do not think this is the case. Because reason is autonomous, it is seemingly able to compete with sentimental inclinations.[22] Kant argues, and I agree, that like the unsympathetic altruist, all people have it in their capacity to rationally isolate themselves and act or strive to reach the moral laws. Change is possible on these grounds.

However, there are reasonable challenges to this idea of change. Consider once again the unsympathetic person. Despite becoming charitable, it is unclear whether her personality changes. It seems as if by not changing her inherent personality, she has not changed at all. However, this change may never be possible, nor does it really matter. This is because personalities of this sort are based on sentimental inclinations. Changing inclinations is near impossible within Kantian understanding, if we take what I have noted above, for there is no objective metaphysical basis by which an appeal to change can be made. This leaves only reason, and reason cannot alter inclinations, but can only exclude them from our maxims. When this unsympathetic character acts charitably by reason, this does not change her personality but does change her internal maxims, and through this can change her behavior. Kant is not concerned with inclinations, which by definition are transitory, but with the rational aspect of humanity. A person may be cold in disposition and yet still appeal to reason to start acting morally by adding this objective law into his or her maxims, and therefore change. Indeed, Kant understands that "cold-blooded goodness is better than a warmth of affection; because it is more reliable."[23]

Certainly, it is unnecessary for a person to feel and to have sympathy feelings, to even be a sympathetic person, in order to be a moral person. Sympathy, which is nothing more than any other sentiment, does not bring about a person's moral status. Nor does the lack of sympathy, as in the case

[22] Kant, *Grounding*, Ak. 448-9 & 452-3.

[23] Immanuel Kant, "Duties Towards Others," in *Lectures on Ethics*, trans. Louis Infield (Indianapolis/Cambridge: Hackett Publishing Company, 1963), 199.

of our unsympathetic altruist, preclude one from achieving morality. To be a moral person, it is unnecessary to say that a person must be sympathetic. Sympathy, as non-objective sentiment, can lead one to both moral and immoral ends. Kant goes so far as to suggest that sympathy inclinations cannot sustain one's morality. Kant notes that a person "benevolent from love, who loves his neighbour from inclination . . . will be charitable, by inclination, to all and sundry; and then, if someone takes advantage of his kind heart, in sheer disgust he will decide from then onwards to give up doing good to others. He has no principle by which to calculate his behaviour. Therefore, the moralist must establish principles, and commend and inculcate benevolence from obligation."[24] The difference, indeed the foundation for morality, is this autonomous rationality and obligation of the categorical imperative.

Sympathy feelings are unnecessary for morality, but Kant allows that morality, practiced through the categorical imperative, may result in sympathy inclinations. "If we do good from duty," Kant says, "it becomes a habit and we ultimately do it from inclination."[25] Such a causal relationship is difficult to establish, for it is not clear that the moral duty necessarily leads to sympathy sentiments. I can only say with certainty that duty and reason are necessary for morality. The unsympathetic person, whether remaining cold in disposition or developing sympathetic inclinations directed with duty, is nevertheless changed from his previous state of uncharitable disinterest to one where duty and morality follow from reason. Reason is the mechanism by which moral development is achieved in either case.

Another criticism derives from a Kantian doubt. This doubt asks whether we are able to know with certainty our maxims, noting that the "insidiousness of the human heart . . . [can deceive] itself concerning its own good and evil attitudes."[26] It would follow that if we were not certain of our maxims, then an implicit bias could prevent us from changing at all. However, it seems as if this does not diminish the rational autonomy per se, only the certainty such capacity gives us. In such a case, we may still appeal to reason to give an approximation of what is moral. While Kant does admit the uncertainty of our maxims, there does seem hope for approximations that we can reasonably deduce by appealing to the idea of maximizing an action to a universal proportion.[27] The unsympathetic character may donate to charity by appealing to this notion despite not knowing whether this is correct.

Furthermore, I find that Kant has strength in his program insofar as we are required to think on our actions, to exercise a "*vigilantia moralis.*"[28] Thinking allows us to ponder morality and what we think we ought to do by appealing to an objective universal law. If we believed ourselves certain, there would be no self-criticism of our maxims. This is the essence of Kant's claims, and it is a strength requiring us not to simply disregard morality despite not being certain, or become arrogant in our certainty. I think it is quite likely that,

[24] Ibid., 193.
[25] Ibid., 195.
[26] Kant, *Religion*, Ak. 38.
[27] Kant, *Grounding*, Ak. 403.
[28] Kant, "Duties to Oneself," 126.

through constant pondering of this law, we can become ever more moral in our actions. Being uncertain does not give us justification for not thinking about morality and acting as we please. Rationality gives us the best option for acting morally.

This paper has sought to answer a crucial question in the Kantian program: what is the mechanism that allows for individuals to morally change? I argued that an individual moral agent's change comes about by appeal to reason and autonomy from inclination. This resistance of non-objective inclination is crucial. While one may understand his duty, it is this autonomous disinterest in sentiments and sensations that is vital for one to achieve this morality. Further, it is the questioning and soul-searching that allows for a constant maintenance of one's maxims, thereby facilitating moral change. Negative change occurs when one rejects the moral law and fails to exclude sentiments from one's maxims. I considered objections to this idea, but ultimately must conclude that change is still possible through an individual's rational autonomy. ❖

Fatalism and Truth at a Time

Chad Marxen

Abstract: In this paper, I will examine an argument for fatalism. I will offer a formalized version of the argument and analyze one of the argument's most controversial assumptions. Then, I will argue that one ought to reject the assumption that propositions *about* the future are true facts *of* the past, even if no one makes reference to such propositions.

Introduction

In this paper, I will present an argument for fatalism. I will argue that the success or failure of the argument is dependent upon the meaning of "true at T." After considering proposed explanations of "true at T" from Nelson Pike and Peter Van Inwagen, I will demonstrate some problems with each explanation and argue that the most plausible interpretation of "true at T" renders the fatalist argument invalid.[1]

The Fatalist Argument

In this argument, let 'E' denote some event:

1. Either E will take place tomorrow or E will not take place tomorrow.
(Assumption)

2. If a proposition about the past is true, then it is now necessary, i.e., inescapable or unpreventable.
(Assumption)

3. If E will take place tomorrow, then yesterday it was true that E will take place in two days.
(Assumption)

4. So, if E will take place tomorrow, then it is now necessary that yesterday it was true that E will take place in two days.
(From 2 and 3)

5. If it is now necessary that yesterday E will take place in two days, then it is now necessary that E will take place tomorrow.
(Assumption)

6. So, if E will take place tomorrow, then it is now necessary that E will take place tomorrow.
(From 4 and 5)

[1] "Future Contingents," *The Stanford Encyclopedia of Philosphy*, last modified June 9, 2001, http://plato.stanford.edu/entries/future-contingents (18 November 2012).

7. Likewise, if E will not take place tomorrow, then it is now necessary that E will not take place tomorrow.
(From 3-5, *mutatis mutandis*)

8. Either it is now necessary that E will take place tomorrow or it is now necessary that E will not take place tomorrow.
(From 1, 6, and 7)

9. Therefore, what will happen tomorrow will happen necessarily, i.e., inescapably, unavoidably.
(From 8)

The conclusion is that what is going to happen tomorrow will happen necessarily. This argument, if successful, demonstrates that every event in the future is fated to occur. Therefore, if anyone doubts the conclusion, then he or she must show which premise is false or which inference is invalid.

The Argument Analyzed

Although the argument seems very plausible, there is an implicit assumption in the argument that needs further examining. The assumption contained within premise 3 is that if a proposition is true, then it is true at all times. But what does it mean to say that a proposition is "true at all times?" Given the previous assumption, the fatalist must clarify what it means to say that a proposition about some time is not only true, but is true at a time that is previous to the time indicated in the given proposition. I will later argue that one ought to reject the assumption that propositions *about* the future are truths *of* the past, even if no one makes reference to such a proposition.

True at T

Although everyone knows what it means to say that a given proposition is true, it is quite different to say that a given proposition is true at some time. Since the fatalist argument requires that there are true propositions *about* the future that are true *in* the past, we need to determine exactly what this means. If the notion of a true proposition *in* the past either is meaningless or is properly understood not to entail that the proposition is a truth *of* the past, then the fatalist argument is unsuccessful. Phrases of the form "true at T" are quite bewildering to several philosophers. Nelson Pike writes, "I am inclined to think that the whole idea of dating the truth-value of a statement in which a date is already assigned to a given event or action, is obscuristic and strange."[2] Peter Van Inwagen states:

> Now I do not think that what we have been offered is a good explanation of the meaning of 'x is true at t' since I don't think this sentence means anything—just as I don't think 'The number twelve is even in Tibet'

[2] Nelson Pike, *God and Timelessness* (New York: Schocken Books, 1970), 71.

> means anything—and thus I don't think anything is or could be an explanation, good or bad, of its meaning.[3]

It seems just as strange to talk about propositions being "true at T" as being "true at L" where "L" denotes some location in the world. What does it mean to say, for example, that 2+2 = 4 in America? Someone may respond, "But I do know what it means to say that 2+2 = 4 in America; it is true in America because it is true everywhere." The failure of this response is that it only pushes the problem back one more level. We can see what it is for a proposition to be true, but what does it mean to say that the proposition is true everywhere? There seems to be a *prima facie* case against the intelligibility of qualifying truth statements either temporally or spatially.

A possible objection, discussed by Peter Van Inwagen, is that temporal qualification of truth statements is intelligible insofar as there are instances of it in regular speech, such that we can in fact make sense of something's being "true at T."[4]At first, the objection seems plausible enough. Suppose I say, "It's raining," and you reply, "That used to be true, but it isn't true anymore." If p denotes "It's raining," then in this instance you can seem to mean "p is true at T_1" and "p is false at T_2." But in fact, what you really mean is that "p at T_1" is true while "p at T_2" is false. In this case, it seems that the previous sentences about raining are denoting different propositions and that this is what happens in ordinary language. When we say things that seem to indicate that a given proposition used to be true but then became false, what we are really saying is that if one asserts the given sentence at T_1, then the corresponding proposition will be true, while if someone asserts the same sentence at T_2, the corresponding proposition would be false. The confusion is a result of the fact that, for the sake of convenience, ordinary language users do not always differentiate between the different propositions expressed by the same statement uttered at different times.

But what if the given translation is implausible and therefore not faithful to what we actually mean when we say the given sentences? Here too, Van Inwagen has an argument to consider.[5] Suppose someone utters the sentence, "The number of players on the basketball team is odd," and someone replies, "It used to be odd, but it isn't anymore." Further, let us stipulate that the number of players on the team used to be 11. The person who replied could mean that the number 11 used to be odd, but that it is not anymore. The person could also mean that the expression "the number of players on the basketball team" used to denote 11, but "the number of players on the basketball team" no longer denotes 11. Since the first translation is absurd, it seems plausible that the second one is correct. If the second translation is correct, then Van Inwagen's translation schema is faithful to our intended meaning in the above sentences.

Although Van Inwagen denies that it makes sense to refer to temporal qualifications of propositions, there are certain sentences that seem to refer to

[3] Peter Van Inwagen, *An Essay on Free Will* (Oxford: Clarendon Press, 1983), 38.
[4] Ibid.
[5] Ibid.

temporally qualified propositions. Consider the following sentence: "George W. Bush will be President of the United States." The sentence is a tough case only if one is an A-theorist concerning time. An A-theorist is one who believes that there exist "A-properties," such as being past, being present, and being future that are not reducible to relations such as later than or earlier than. An A-theorist further believes that there is an objective present that is constantly changing whereas one who is not an A-theorist does not believe that there is an objective present. If an A-theorist is correct concerning the previous claims, then sentences about the future such as "George W. Bush will be President of the U.S." refer to propositions that used to be true, but at the present point in time are false—in which case there are intelligible cases of temporally qualified statements. If the A-theory of time is true, it seems that the fatalist argument can succeed. The fatalist can say that there are truths *in* the past *about* the future that necessitate the actions in the future.

The objection from the A-theorist is formidable, but there seems to be an adequate response to it. In order for the fatalist argument to succeed, propositions *about* the future must be temporal. But since the propositions are without change logically prior to becoming false it seems that the propositions are atemporal logically prior to their losing the property of being true and gaining the property of being false. For this reason, it does not make sense to talk about the propositions at a given time in the past since the propositions are atemporal logically prior to becoming false. Therefore, the fatalist assumption that propositions *about* the future are *in* the past is false. If the truth values of propositions about the future change, then contrary to Van Inwagen, it seems to make sense to say that some propositions used to be true. In the following section of the paper, I will consider different accounts of what it might mean to say that a proposition used to be true. Before I do so, a few more comments regarding the objection from the A-theorist are pertinent.

If the previous reply is unsuccessful and propositions *about* the future are temporal, then what follows? It seems that the proposition, "George W. Bush will be President of the U.S." is fated to occur. But just because it is fated to occur, it does not follow that the event is fated to occur at any particular time. It may be the case that future events are fated to occur even though the exact ordering of the events is not fated. For example, one might be fated to marry a particular person, have a particular job, and live in a particular place although the ordering of these events remains in one's control. It is also important to recall that the previous objection presupposes an A-theory of time, so if that theory is false, then the given objection is unsuccessful. Additionally, since most A-theorists are presentists, they could become non-A-theorists in light of Joshua Rasmussen's recent argument that a presentist need not be an A-theorist.[6]

Since it is not at all clear what "true at T" means, one must consider some proposed definitions of the statement "true at T," where "T" is a time in the past. The first definition comes from Nelson Pike. He suggests that we

[6] Joshua Rasmussen, "Presentists May Say Goodbye to A-properties," *Analysis* 72, no. 2 (2012), 270.

translate the sentence, "p is true at T_1" as "p at T_1" is true.[7] Let's consider an example. Suppose someone says, "'Jones runs', is true at T_2." We can translate that sentence as "'Jones runs at T_2' is true." Suppose someone replies, "The statement 'Jones runs at T_2 is true' is true at T_1." Now the sentence becomes "Jones runs at T_2 at T_1 is true." The problem here is that the original proposition "Jones runs" is indexed to a specific time, T_2, and the new proposition that "Jones runs at T_2 at T_1" is an indexed proposition of the *original* indexed proposition. If Pike's definition is correct, then the fatalist argument will be unsuccessful since we will not be able to index a proposition to a given time if the proposition is *already* indexed to a different time. If fatalists merely want to index propositions that are not already indexed to a given time, this will do them no good because the stated proposition will merely be *about* the future, but it will not be a fact *of* the past.

Van Inwagen has proposed an alternative to Pike's definition of "true at T." Van Inwagen states, "to make sense of this idea [viz., "true at T"], it would be sufficient to make sense of the open sentence: (The proposition) x is true at (the moment) t." Both he and Gilbert Ryle propose, "If someone were to assert x and nothing else at t, then what he asserted at t would be true."[8] There is a problem with the previous definition.[9] Consider a state of affairs in which no propositions have been asserted in 10,000,000 BC or earlier. If the notion of "true at T" makes sense, then if we let "p" denote the proposition that no propositions have been asserted in 10,000,000 BC or earlier, it would seem that p is true in 10,000,000 BC. But if someone asserted p in 10,000,000 BC, then p would turn out to be false by the very act of asserting it. Therefore, if p is true in 10,000,000 BC, then according to the definition, if someone asserted p in 10,000,000 BC, what he asserted would be true. However, if he asserted p, then what he asserted would be false. P would be false because a proposition would have been asserted in 10,000,000 BC or earlier. Therefore, according to the definition, what he asserted would be both true and false.

Although Van Inwagen's definition has a problem, it seems to be the most intuitive definition of what it means to say that something is "true at T". Although the definition produces a paradox, classic logic is similarly paradoxical, as illustrated by the liar paradox. In this paradox, consider the sentence, "this sentence is false." If it is true, then it is false, and if it is false, then it is true. Therefore, it may be unreasonable to disregard the fatalist definition because of the technical problems just as it would be unreasonable to dismiss classical logic because of the liar paradox. Additionally, since Van Inwagen's definition seems to be the most plausible definition of "true at T," it is not clear what else it could mean. Therefore, the previous definition seems to be the only definition that can give the fatalist argument a chance for success.

It is now appropriate to consider the second and third premises with the Van Inwagen and Ryle translation schema:

---

[7] Pike, 68.
[8] Van Inwagen, 37.
[9] Ibid., 228-9.

2. If a proposition about the past is true, then it is now necessary, i.e., inescapable or unpreventable. (Assumption)

3. If E is going to take place tomorrow, then it is true that yesterday it was true that E would take place in two days. (Assumption)

Since we now have a definition for "true at T," let us look over premises 3 through 5 with the given translations.

3*. If E is going to take place tomorrow, then it is true that if someone had asserted yesterday that E would take place in two days and nothing else, then what he asserted would be true. (Assumption)

4*. If E is going to take place tomorrow, then it is now necessary that if someone had asserted yesterday that E would take place in two days and nothing else, then what he asserted would be true. (Follows from 2 and 3)

5*. If it is now necessary that if someone had asserted yesterday that E would take place in two days and nothing else, then what he asserted would be true; then it is now necessary that E is going to take place tomorrow. (Assumption)

It seems that the inference from 2 and 3* to 4* is invalid. While it is the case that the consequent of 3* is *about* the past in the sense that it is about what *could have* happened in the past (i.e., the past of a different possible world), it is not *about* the past in the sense that it is a fact *of* the past about *what actually* happened in the past. Since the consequent of 3* does not fall under the scope of premise 2, and because the scope of premise 2 only includes the actual past, the inference from 2 and 3* to 4* is now fallacious.

Additionally, even if one were to allow the previous inference, premise 5 becomes very implausible in light of the translation schema. The inference essentially says that if the consequent of the previous premise—which is a conditional—is necessarily true, then it is necessary that the future event will occur. But clearly the previous inference is problematic. The inference commits a modal fallacy by confusing the necessity of the *inference* with the necessity of the *consequent*. Since it would only be the case that the *inference itself* is necessary, unless the antecedent were *necessarily true*, it does not follow that the consequent is *necessarily true*. And if it does not follow that the consequent is necessarily true, then it does not follow that it is now necessary that E is going to take place tomorrow. Since the inference from 2 and 3* to 4* is fallacious, and 5* is false, the fatalist argument is unsuccessful.

Conclusion

In this paper, I have argued that the fatalist argument hinges on the assumption that there are truths *about* the future that are facts *of* the past, even if no one makes reference to a proposition about the future. Since we have no reason to accept this assumption and have reasons to reject it, the fatalist argument is unsuccessful. I have not considered whether or not there are truths *about* the future that are facts *of* the past in virtue of a person referencing the future through asserting, uttering, or believing a proposition about the future; so it may be the case that the action of referencing a proposition *about* the future becomes a fact *of* the past and thereby *necessitates* the future event. Since dealing with this problem is not within the scope of my paper, I leave it to others to decide whether or not referencing the future makes any future event necessarily occur. ❖

Indiscernibles and Plato's Forms vs. Parmenides

Jenny Carmichael

Abstract: In *Parmenides*, the young Socrates defends several candidate forms against Parmenides, who makes five objections: the objection of forms of common things, the question of the part vs. the whole, the third man argument, infinite regress, and the greatest difficulty problem. I define forms in terms of Leibniz's Principle of the Identity of Indiscernibles (PII) in an attempt to overcome Parmenides' opposition. I show that the main force in Parmenides' objections consists of absurdities that emerge in relations between forms and particulars: absurdities that are avoided if the form and its instantiation in the particular are identical.

Indiscernibles and Plato's Forms

Plato's *Parmenides* provides an informative account of forms as well as a rigorous series of arguments against them. Socrates is a young man in this dialogue and has no well-articulated account of the forms, so there is an aporetic element to the dialogue concerning the answers to Parmenides' objections that persists with special regards to the third man and the greatest difficulty objections. I suggest the Principle of the Identity of Indiscernibles (PII) as an additional criterion for the general account of forms, which helps overcome Parmenides' objections while remaining consistent to Plato's theory of forms in the dialogue. When applied to aspects of the inchoate theory, the PII resolves Parmenides' five objections to Socrates' account by defining the way particulars instantiate forms as an identity between the form and its instantiation in the particular.

The format of the paper is as follows: I first explain how and why the PII should be applied to the theory of forms, and entertain four objections to the explanation. I then go on to explain and respond to each of Parmenides' five objections in the following order: forms of common things, the part vs. the whole, the third man, infinite regress, and the greatest difficulty problem.

The PII is Leibniz's claim that if x and y have indiscernible sets of properties[1], then x and y are identical.[2] The reason for applying it to the theory of forms in the Parmenides is because of the following account of forms that Socrates initially gives to Parmenides:

> But tell me this: don't you acknowledge that there is a form, itself by itself, of likeness, and another form, opposite to

[1] The use of the word "properties" here is not meant to be an ontological claim, but as a neutral way to describe instantiations of forms in particulars.

[2] G. W. Leibniz, *Philosophical Essays*, trans. Roger Ariew and Daniel Garber (Indianapolis: Hackett Publishing, 1989), 41-2.

> this, which is what unlike is? Don't you and I and the other things we call 'many' get a share of those two entities? And don't things that get a share of likeness come to be like in that way and to the extent that they get a share, whereas things that get a share of unlikeness come to be unlike, and things that get a share of both come to be both?[3]

If the ability of particulars to instantiate forms in degrees is taken to imply that an aspect of the particular has qualities that are indiscernible from the quality of the form, Parmenides' criticisms of forms can be avoided. With this interpretation, the degree to which a particular thing instantiates the form is the degree to which the particular shares identity with the form. For example, if a set of three objects was partaking in the form of oddness, then what can be said of the set regarding oddness is indiscernible from the form of odd. If the set instantiates something that is discernible from oddness (in this case, evenness is the only alternative), then there is no reason to say that it instantiates that form, if, as Socrates claims, it comes to be such a way because of the form. If it were taking part in the form, there would be something about it that can be said of the form. Since a form is one trait,[4] or "itself by itself,"[5] if the form is said of something, the thing must have the one trait. Otherwise, the form is not present.

One may object that no particular physical set of objects can be considered a paradigm case of something like oddness, since many sets of many different numbers equally qualify, and therefore, is discernible from the form of oddness. But this can instead be interpreted as the form of oddness being instantiated amongst instantiations of a variety of other forms. If they are said to be odd at all, something about them must be consistent with the pure notion of oddness, otherwise describing them as such is just false. If a thing were to get a share of a form as Socrates claims, it would be strange if the share were missing the essential trait of the form. It seems impossible to find an example of a numerical set that is partially odd—and if one could do it, such a set would not instantiate the form of oddness.

Further, one may be reluctant to accept the explanation because it is an uncomfortable notion that a specific instantiation with regard to a particular object is identical with a form, since not every instance of oddness looks like this one. But, as previously stated, the other traits of the set, outside of this particular form, detract from the trait that is the form of oddness. Other things can be said about the set, such as it being a set of tables, whereas nothing can

---

[3] Plato, "Parmenides," in *Complete Works*, trans. Mary Louise Gill and Paul Ryan (Indianapolis: Hackett Publishing, 1997), 129a.

[4] I will use the neutral term 'trait' to describe the form's relation to the particular. Plato's Socrates acknowledges a similar terminological problem (Phaedo 100d): "[N]othing makes a thing beautiful but the presence of, or the sharing in, or however you may describe its relation to that beauty we mentioned, for I will not insist on the precise nature of the relationship except that all beautiful things are beautiful by beauty." The difference is that I insist that the precise nature of the relationship is identity.

[5] Plato, 129a.

be said of the form of oddness except that it is odd. Given that the qualities in the instantiation of the form are indiscernible from the quality of the form itself, then by the PII, the form and the instantiation of it must be identical. This does not mean that the entire particular is identical with the form; it only means that if an object instantiates a form, then there is a trait within the total set of the object's traits that is identical with the form.

Here, it may be argued that Plato would never accept this way of defining forms in consideration of Gregory Vlastos' nonidentity assumption, by which a particular object, x, cannot be identical to the F-ness by which it instantiates the form F.[6] It may be objected further that, in 130b, Socrates agrees with Parmenides' statement that forms are distinguished as separate from the traits that instantiate them, specifically stating that "likeness itself is something, separate from the likeness we have."[7] One may also say that he blatantly rejects the PII as applied to forms: "On the other hand, if it were the same as another, it would be that thing, and not itself. So in this way, too, it would not be just what it is—one—but would be different from one. Therefore, it won't be the same as another or different from itself."[8]

In response to Vlastos's objection and to the objection concerning the first passage, it can still be true that likeness itself is separate from something partaking of it if the form is only identical with its instantiation in the object. The form is distinguished from the particular with regard to the PII because they do not have identical sets of properties.[9] Consider again the set of three tables. The tables are brown, hard, and odd. The odd itself is only odd. Therefore, although the odd is identical with the trait of oddness that the tables have, it is not identical with the set of tables. The only identity is between the trait that is the form in the object, and the form itself. This is not sufficient for making the object and the form identical.

As for the objection concerning the second passage, when Parmenides states that the form of the one is not the same as another thing, Parmenides can be interpreted as speaking of the other thing in terms of other forms—not of other objects. If there were two identical forms of the one, there must be something to distinguish them, so one of them would not be identical with the one. Thus, he reaches the conclusion that there is only one form of each trait for which there is a form. Further, this passage is fully consistent with the PII in regard to forms, because if anything is exactly the same as something else, it follows that there is an identity relation. Plato rejects that two identical yet numerically distinct things can exist, so if there is identity, there must be

[6] Gregory Vlastos, "The Third Man Argument in the Parmenides," *Philosophical Review* 63 (1954): 325.

[7] Plato, 131a-e.

[8] Ibid.

[9] I do not claim here that forms have sets of properties; instead the form itself is a property. So if something has a property other than the instantiation of the specific form, the conditions for identity laid out by the PII are not met.

numerical sameness, as is the case with the form and its particulars. This hinges on the concept of multilocation that is implicit with the PII.[10]

Forms of Common Things

The first of Parmenides' objections is the question of whether there are forms of common things, e.g., "hair, mud, and dirt," which puts Socrates in a quandary, having no way of deciding what sorts of things exist as forms. But this difficulty no longer stands with the PII explanation, unless the common objects themselves were forms. The possibility for the form to be instantiated in a particular is necessary, otherwise there would be no reason for postulating it as a form. So under this explanation, the form must be able to exist multilocally and retain its identity, because it is assumed that many objects are partaking of it, and that each object is in a different location. The reason that things like mud and hair are unable to fit this definition is that the identity amongst things partaking of muddiness, and muddiness itself, exists only linguistically (or perhaps as an Aristotelian form). It is true that two samples of mud both have traits of muddiness (e.g., wetness, graininess), but there is nothing about the notion of muddiness that shares identity with the particulars. For example, if there were two molecularly indiscernible samples of mud, it would be incredibly anomalous, and an unnecessary qualification for both to be considered mud. The term muddiness would accurately describe both samples, but the recognition that connects the two would be based on traits that are not necessarily indiscernible. Further, the indiscernibility of mud samples and muddiness, if it were possible, would depend on a molecular structure, which poses a problem: namely, the only thing that would be able to instantiate such a form would be something with exactly the same molecular structure. This would limit the sharing in the form to things with identical structures, and they would no longer be differentiable from the form itself, due to the PII. Furthermore, as Samuel Rickless points out, mud is limited to a spatiotemporal definition, meaning that it can only be explained in physical terms.[11] Rickless has different reasons for rejecting such a form on these grounds, but another consequence of this theory is that a spatiotemporal object would have to be multilocated to fit the PII explanation, and this would certainly be a problem.[12] With the PII explanation in addition to Rickless' notion, it is clear that a form must have the ability to exist multilocally, and that Parmenides' common objects do not have this trait.

---

[10] The term multilocation does not assume spatiotemporal traits here. A wall, instantiating the form of blueness, for example, gives us two locations of the form of blueness: the form of blueness within the wall, and the form of blueness itself. The wall gives us an instance of the form pertaining to a spatiotemporal object, and the object's spatiotemporal traits are not a result of the form. Given that the form cannot be spatiotemporal, this would not be possible. Cf. Dean W. Zimmerman, "Distinct Indiscernibles and the Bundle Theory," *Mind* 106, no. 422, (1997) 305-9.

[11] Samuel Rickless, *Plato's Forms in Transition* (Cambridge: Cambridge University Press, 2007).

[12] Max Black, "The Identity of Indiscernibles," *Mind* 61, no. 242 (1952): 153-64.

Part vs. Whole

Parmenides' next objection is the difficulty of the part versus the whole.[13] The force of this objection comes from the absurdity of a thing instantiating a part of the form, and thus having traits from the form that are opposite of the form itself; in other words, a part of the large that is smaller than the large itself. Socrates is also hesitant to admit that an entire form is instantiated in a particular, but it is clear how the PII explanation of forms and particulars counters this objection, since it becomes essential that each particular that instantiates the form instantiates the entire form. If the form itself is identical with its instantiation, then the instantiation cannot be a part of the form, for then the two would not be identical. Take, for example, whiteness. When two pieces of paper are white, assuming the shades are indiscernible, the whiteness of each paper is identical to the form of white. It is not said that the whiteness of one paper is identical with one part of whiteness and the other paper identical with another part. Instead, each paper is simply said to be white; its whiteness is the form of whiteness being instantiated in the particular object, i.e., the pieces of paper. The form and its instantiation are indiscernible, thus identical. If they are identical, it is impossible that one be a part and the other a whole, so the instantiation is necessarily one with the form itself.

The Third Man

The third man problem is a rather challenging problem for Socrates on which at least two other problems rest. The third man problem is the regress that comes about when a form, having all the properties of which it is a form, is compared to its occurrence in a particular, thus generating a comparison which relies on another form, which is again compared to the first form by another, and so on ad infinitum.[14] Rickless outlines the problem nicely: if there is a set of large things, {a, b, c}, each taking part in the form of the large, and the large itself is large (thus taking part of itself), the set, P1, will look like this: {a, b, c, L1}, for which L1 is the form. But if each thing in the set is in the set because of instantiating the form of large, then there must be another form, L2, by which the members of the set are able to instantiate the form of large. This is because, according to Rickless, a form cannot instantiate itself.[15] If this is true, it follows that the L2 is necessary for L1 to be large. It is apparent, then, how the number of forms of the same thing go on to infinity, for L3, L4, L5, and so on, can be added to the set creating a demand for another form infinitely many times.[16]

So, applying the PII to the forms and their instantiations rejects Vlastos' self-predication assumption and Rickless' non-self-partaking auxiliary. If a form has the trait of which it is a form, it is impossible with the PII for it

[13] Plato, 131a-e.
[14] Ibid., 132a-133e.
[15] Rickless, 71.
[16] Ibid., 67-8.

not to instantiate itself; for a thing to be able to instantiate the form, the form must not only be, but also have a trait with which a trait of a particular can be identical. If so, it becomes almost meaningless to compare the form to the particular in the first place, since the only comparison between the two is the identity between them. Thus, if a building is large by instantiating the form of largeness, and the form of largeness is large in virtue of itself, it seems that the form itself must be more large than the building, since it would be absurd for something to be more like the form than the form to be like itself. But with the PII, it cannot be supposed that the form of large is larger than the largeness of the building, since the building, insofar as it is large, is identical with the form, and the form cannot be larger than itself.

Infinite Regress

There is a similar answer to the objection of the infinite regress problem.[17] This is the problem that there can be nothing like a form, nor can there be a form like anything else, since things have likeness due to instantiating the same form, and so if a form and a thing have likeness, another form must supply it. This objection is answered in the same way as the third man argument. The similarity between two particulars, insofar as it is indiscernible, is identical, and if this similarity is a form, there is no need for another form to exist in order to contain the identity.

The Greatest Difficulty

Parmenides' final objection is the greatest difficulty objection,[18] or the Access Problem[19]—namely, the question of how humans are able to know forms when our knowledge is restricted to "things that belong to us" while a form, by being "itself by itself,"[20] necessarily does not belong to us, and that proving the existence of forms is nearly impossible for anyone who is not "widely experienced and not ungifted."[21] Thus, if forms are non-spatiotemporal, and everything of which we have knowledge is known through the senses (thus belonging to us), then a theory of forms seems to be speculative, without empirical evidence in regards to mathematical platonism. Benecerraf points out that there must be access to non-spatiotemporal entities, given that we have knowledge of mathematics, which is non-spatiotemporal.[22] To preserve the plausibility of forms, there must be some relation between these non-spatiotemporal objects and our spatiotemporal world that we can recognize. Thus, forms must be understood similarly to the way mathematics

[17] Plato, 132e-133a.
[18] Ibid., 133b-135c.
[19] This term is introduced in Paul Benacerraf, "Mathematical Truth," *The Journal of Philosophy* 70 (1973), 666-8: the idea that humans must have access to some mathematical truths (though not necessarily all), since we are able to posit true ideas about them.
[20] Plato, 133c-d.
[21] Ibid., 133b.
[22] Benacerraf, 666-8.

is understood if, like mathematics, they are to have any reality that is not contingent on human thought.

The PII relates forms to spatiotemporal objects by means of identity. The form of a non-spatiotemporal trait is identical to the same non-spatiotemporal trait of the spatiotemporal particular that instantiates it. Parmenides notes that when we encounter a form, the encounter is not with the form alone, but instead with its instantiation in some particular thing.[23] If there is an identity in such a circumstance, then Parmenides must admit that we know the form this way. An object instantiating some form has a trait that is indiscernible from the form, and if we accept the PII and posit an identity, then by claiming that we can know the forms as they pertain to particulars, we can know the forms. With a mathematical theorem for example, it would be incorrect to say that the knowledge necessary for writing and understanding a theorem is only partial, and limited by our ability to sense only spatiotemporal objects. Instead, if the theorem is consistent with mathematics in general, the working of the theorem and the mathematical truth are indiscernible.

For the PII explanation of the relation between forms and particulars to be correct, it must be possible for spatiotemporal objects to have non-spatiotemporal traits. If they do not, then it is impossible for particulars to instantiate forms. Under such a view, the theory of forms is both ontologically and epistemologically useless; it makes no difference whether or not we can know forms if it is impossible that they are instantiated in the world of sensibles.

The PII may seem to make forms of things that are almost unanimously rejected from formhood, e.g., colors, largeness, justice, and so on. But this does not imply that all sets of indiscernible things are candidates for formhood. This series of arguments is not sufficient for applying the PII to all Plato's mentions of forms, limited as it is to the Parmenides. It fails to take into account other factors that both include and rule out other candidate forms. So, although it can be contested whether the PII provides a plausible explanation of forms, its application with regard to Parmenides' objections is a good predictor of how consistent it might be with candidate forms from other dialogues.[24] ❖

[23] Plato, 134a.

[24] I am grateful to Debra Nails and Erik Jensen for helpful comments, and to Dustin Brown for a thorough proofreading.

Instrumentalism and Poetic Thinking:

A Critique of Dewey's Logic of Thought

Mark McGinn

Abstract: This paper offers a critique of the instrumental logic of thought found in the middle period of Dewey's philosophy. His instrumentalism requires that thought serves to effect a physical alteration in the conditions of experience through an experimental act, the results of which retrospectively determine the legitimacy of thought. But missing from his account, I argue, is an explanation of the significant alteration of experience brought about by more aesthetic forms of philosophical thinking, which do not aim to effect any kind of physical alteration. I therefore propose that "poetic thinking" be invoked as a necessary supplement to instrumental thinking.

Introduction

One of John Dewey's most important innovations as a philosopher was his introduction of a new logic of thought, as laid out in his *Essays in Experimental Logic.*[1] Past schools of thought, he contends in this work, have placed too great an emphasis on the processes of reflection and inquiry themselves, without considering the non-reflective context in which thought is situated. If this context is recognized in its full import, it becomes clear that an account of the temporal development of experience must figure largely in any adequate logic of thought.[2] Reflection and inquiry are then found to occupy an intermediate and mediating position in the development of experience; they are found to be instrumental, meaning they serve to effect a physical alteration in the extant conditions of experience.[3] Dewey is convinced that this framework provides for all the possibilities of legitimate philosophical thinking.

This essay offers a critique of this notion. In Section II, I begin by providing a brief exposition of Dewey's account of the temporal development of experience and its constitutive moments—pre-reflective experience, reflective experience, and the post-reflective situation—in order to set the stage for what follows. In Section III, I proceed to the critique itself, where I suggest that Dewey's theory is legitimate (and even groundbreaking) within its own limits, but constricting to a survey of the broader possibilities of philosophical thought when taken to exhaust those possibilities. His instrumentalism requires that thought produce a physical alteration of the conditions of experience through an experimental act, the

[1] John Dewey, *Essays in Experimental Logic* (Chicago: The University of Chicago Press, 1916).
[2] Ibid., 1.
[3] Ibid.

results of which retrospectively determine the legitimacy of thought. Missing from his account, however, is an explanation of the significant alteration of experience brought about by more aesthetic forms of thinking, which do not aeffect—nor intend to effect—any kind of physical alteration. Therefore, I propose that "poetic thinking" be invoked as a necessary supplement to instrumental thinking. In this way, one avoids the difficulties that appear when the latter is taken to account for all forms of philosophical thought.

Exposition

The first moment of the temporal development of experience, which Dewey calls by several names, is pre-reflective experience.[4] This moment comprises what some may refer to as a "knowledge" experience (as when we speak of "acquaintance knowledge," "immediate knowledge," or skill and habit) although Dewey is convinced that this type of experience cannot properly be called a knowledge experience without perverting both of these terms. He explains that it may be guided by knowledge resulting from previous inquiries, and may even contain an element of thinking, but not in such a way that they structure the situation and lend it a pervading quality.[5] For a concrete example of this pre-reflective mode, consider the difference between, on the one hand, an experience of drinking water where the perception of water is peripheral to the action itself, and on the other hand, an experience where knowledge of the constitution of water is the controlling interest. In the former case, water is experienced in an unreflective, practical way, as something encountered in the midst of everyday concerns and projects; in the latter case, it is experienced primarily as an object of perceptual and cognitional apprehension, that is, as an object of knowledge.[6]

Philosophers, however, have a natural tendency to distort the unreflective mode of experience by attributing qualities to it that in fact belong to the cognitional mode. Such a misconception, Dewey claims, results in a falsification of the way we phenomenologically experience the world. Because philosophers do not think about knowledge except through the lens of reflective experience, they are predisposed to regard all experiences as if they were of the same sort as reflection.[7] The result is that they inadvertently attribute qualities to the whole that in fact are peculiar to one of its parts. Thus, whereas things are present to us most of the time in experiences of desire and antipathy, pleasure and pain, reluctance, indifference, and reminiscence, the professional philosopher tends either to think of such things as objects of conscious knowledge or to disregard them altogether.[8] While this inclination is pardonable insofar as it arises in homage to the value of thought and

[4] As Dewey employs no fixed vocabulary in his writing, I have resorted to coining two terms for the sake of simplicity: "pre-reflective experience" and "the post-reflective situation." The reader should bear in mind that Dewey does not use these phrases.
[5] Dewey, 2.
[6] Ibid.
[7] Ibid., 2-3.
[8] Ibid., 3.

reflection—a value that should by no means be underestimated—it is nevertheless damaging to critical philosophical investigation. Any inquiry that proceeds from this misconception is likely to be misguided from the start.[9]

Another important aspect of experience (whether pre-reflective or reflective) is that it is set in a non-cognitive context that holds in place a vast network of interrelations that inflect the focal object with meaning and significance. In other words, experience is structured by an internal organization by virtue of which the relations that comprise it "hang together" and are imbued with a saturating quality.[10] Consider, for example, how the focal point of experience—right now, the words of this essay—are surrounded by the page, the room, the building, the town, and so on, each successive level becoming increasingly indistinct, and how, in addition to these spatial horizons, one's projects, habits, interests, and past stretch backward and forward in time, such that one interprets the environing world by their lights.[11] We find that experience is ensconced in sundry horizons, both spatial and temporal, which shade off indefinitely into the fringe of awareness and beyond, infusing its focal object with significance. Although as philosophers we tend to think of such internal organization as the outcome of conscious thought processes, it is actually, according to Dewey, constituted largely by experiences that are non-reflective in character. This immense and active nexus of varied and interrelated elements is precisely what he means when he speaks of "experience."

Granted, then, that the greater part of experience is unreflective in character or at least constituted by unreflective factors, the task remains of describing the characteristics and purpose of *reflective experience*, the second moment in the temporal development of experience. Dewey claims that reflection first arises when the factors that comprise the empirical situation just described come into conflict; it is brought to the forefront of experience when something goes wrong, when friction and discordance disrupt our habitual, unreflective engagement with the environing world.[12] He emphasizes that such incompatibility of factors is not merely structural or static, as would seem to be the case if we supposed that our previous reflective inquiries had simply been inadequate in ordering experience. Instead, the factors of the empirical situation are incompatible in an "active and progressive" sense, as living hindrances to projects that extend beyond our intellectual concerns.[13]

Since incompatibility of factors obstructs some active interest of the knower, a particular purpose is set for thought in each case. The solution of this conflict (or, more accurately, the *re*-solution back into non-conflicted experience) must remain faithful to the existing conditions of the empirical situation. Only in relation to these may thought take its proper orientation. For example, if one ignores these conditions in formulating a solution, or evades them by escaping into imagination and fantasy, the conflict is either further agitated or remains unsettled (if it does not resolve itself). Conversely,

[9] Ibid.
[10] Ibid., 4.
[11] Ibid., 6.
[12] Ibid., 10.
[13] Ibid.

if the problems one attends to do not arise from any real conflict—if they are problems purely of thought, divorced from experience—they should not be regarded as problems at all. Also, as long as the results of thought are not put into effect by being tested in action (as long as they do not serve as an instrument in effecting change in the empirical situation), the so-called knowledge gained lies fallow and thus does not qualify as knowledge at all. In order to become effective knowledge, it must translate into experimental action and thereby generate another non-reflective situation within which conflicts may again arise and new problems for reflection are set. I will refer to this second non-reflective situation as the post-reflective situation.

The post-reflective situation differs from the pre-reflective situation in three important ways. First (and foremost, for Dewey), the actual physical conditions of the situation are altered.[14] For example, if the factors of a particular conflict are thirst and the presence of unclean water, then the purification of that water—say, by boiling it—will result in a physical modification of the factors of experience: the cleansing of water and the quenching of thirst. The second way in which the situation is altered is through the changed character of the agent herself. With each successful operation of thought, she acquires new habits and skills, such as those involved in purifying water, that will be of use in future conflicts.[15] Finally, the post-reflective situation differs from the pre-reflective situation in that the factors involved in the experience accrue meaning and significance for the agent.[16] To continue with the example: Never again will she look at unclean water in the same way, as something undrinkable, and the tools she used to purify it will accrue meanings that were not previously there, presenting themselves as things that are useful for cleansing water.

As we will see, Dewey accords a certain priority to the first of these changes (physical alteration), while the latter two are regarded as secondary and posterior. I wish to call this priority into question. All the same, it should be kept in mind that this essay, whatever critique it may make of Dewey's instrumentalism, remains within the structure established by his distinctive notion of experience. This I leave untouched.

Critique

Dewey's logic of thought, as he himself says, is instrumental. Thought is measured by the extent to which it serves as an instrument in effecting a physical change in the environment.[17] However, we are justified in asking whether this should be the only goal that thought sets for itself and more specifically, whether it should be the sole end toward which philosophical thinking aims. Might the possibility exist of a mode of thought that is interested

[14] Ibid., 31.
[15] Ibid., 16.
[16] Ibid.
[17] Dewey is quite clear about this point: "The reorganization, the modification, effected by thinking is … a *physical* one. Thinking ends in experiment and experiment is an *actual* alteration of a physically antecedent situation." Ibid., 31 (first italics mine).

in more than effects and results, a mode of thought that moves on more subtle levels?

Dewey suggests this possibility when he says that aside from being instrumental to gaining control of situations, thought may also serve to "enrich . . . the immediate significance of subsequent experiences," hinting at a poetic mode of thought whose purpose is to enhance the meaning and quality of life, though not in any kind of physical way. "And it may well be," he continues, "that this by-product, this gift of the gods, is incomparably more valuable for living a life than is the primary and intended result of control."[18] He seems to soon forget this possibility, or else to relegate it to the realm of art and literature, when a few pages later he reiterates that thought must produce a physical modification of the environment. We are led to conclude that such poetic thinking lies outside the proper domain of philosophy, that it is merely a "by-product" of "the primary and intended result of control." But should these by-products, which even he suggests are inimitably valuable, be regarded as incidental to thought? Perhaps there is a mode of thinking whose express purpose is to enrich and deepen the significance of experience, to forge value. Perhaps we may reasonably posit, in addition to instrumental thinking, what may be called poetic thinking, a mode of thought that aims not at the physical alteration of the conditions of experience but rather at the *significant alteration* of experience.[19] Before moving on to expound the purpose and characteristics of poetic thinking, it will be beneficial to illustrate in what ways instrumental thinking comes up short when it is taken to exhaust the possibilities of philosophical thought, as well as the dangers inherent in taking up such a view. The writings of Randolph Bourne and Martin Heidegger will aid us in this task.

At one time a disciple of Dewey, Bourne poses questions similar to those we have considered. His chief criticism is that Dewey's instrumentalism, which concerns itself primarily with technique and expediency, is apt to come with a loss of "poetic vision."[20] Instead of striving to create new values and open new horizons of thought, it settles with the goal of adaption to a pre-existing environment. "The defect of any philosophy of 'adaption' or 'adjustment,'" Bourne writes, "is that there is no provision for thought getting beyond itself. If your ideal is to be in adjustment to your situation, in radiant co-operation with reality, then your success is likely to be that and no more."[21] Thus, for Bourne, Dewey's instrumentalism suffers from a devastating lack: It fails to account for the visionary side of thought, which he believes must "constantly

[18] Ibid., 17-8.

[19] I do not mean for this distinction to suggest that philosophers must be either instrumental or poetic thinkers; the work of most philosophers contains both instrumental and poetic elements. Nevertheless, some tend so far in one or the other direction that they serve as paradigm examples of one particular mode. Dewey and Marx, for example, who both fervently advocated social and political reform, are typically instrumental philosophers. Nietzsche, Heidegger, and Wittgenstein, who aimed primarily at transforming the understanding of their readers, are characteristic examples of poetic philosophers. (These lists are by no means exhaustive, of course.)

[20] Randolph Bourne, *War and the Intellectuals: Collected Essays, 1915-1919*, ed. Carl Resek (Indianapolis: Hackett Publishing Company, 1999), 61.

[21] Ibid.

outshoot technique."[22] It stops short of what is possible as such, and limits itself to the circumscribed possibilities of a given set of conditions. By thus being assigned determinate limits, the field of potential human endeavor and thought is restricted.

Martin Heidegger points to a different, though related, shortcoming of instrumental thinking. While he acknowledges that it is justified and even necessary for living a life, he claims that instrumental thought (which he refers to as "calculative thinking") leads to thoughtlessness when carried to an extreme. In "always reckon[ing] with conditions that are given," it is prone not only to overlook what is possible as such but to lose itself in a frenzied ordering of the actual. "Calculative thinking races from one prospect to the next," he writes. It "never stops, never collects itself."[23] Whereas Bourne is concerned that instrumental-calculative thinking restricts philosophy's range of possibilities and thus limits it, for Heidegger the risk is that we may come to expect too much of philosophy by demanding that it serve to manipulate the environing world in some way, and thereby overlook its less apparent effects.[24] He suggests that much philosophical thought moves on more subtle levels, such as attuning us with aspects of experience that normally pass unnoticed, evoking new modes of comportment toward the world, venturing into unexplored realms of thought. Philosophical thinking may also take the form of "self-meditation," in which case thinking is directed back upon itself in an effort to see through its own determinations and historical constitution.[25] In demanding of such thinking that it produce physical changes in the world, Heidegger claims, we impose a foreign standard upon it. Its effects may be only mediate—but they are no less decisive.

One of my chief worries with Dewey's instrumentalism is that, when adopted by others, it may become so forward–looking that it fails to recognize its own determinations. In order to be effective, as Dewey himself says, thought must bleed into experience, so to speak. But experience, we should remember, also bleeds into thought. We would be remiss to assume that one may adopt a wholly unbiased standpoint from which to philosophize; we bring much with us into reflection that is unreflective, assumptions so deeply embedded in our culture and concepts as to be imperceptible. These historical determinations shape and guide our inquiries to a great extent. Thus the risk of instrumental thinking, which ever keeps its sights on tangible results, on the future and progress, is that it may become so forward-looking as to let its determinations go unnoticed. One of the tasks of a more meditative, poetic mode of thinking, the kind Heidegger advocates, is to supplement and offset the blind spots inherent in instrumental-calculative thinking (necessary as it is) in order to

---

[22] Ibid.

[23] Martin Heidegger, "Memorial Address," in *Discourse on Thinking*, trans. John M. Anderson and E. Hans Freund (New York: Harper Perennial, 1966), 46.

[24] Martin Heidegger, *Basic Questions of Philosophy*, trans. Richard Rojcewicz and Andre Schuwer (Bloomington: Indiana University Press, 1994), 5.

[25] Martin Heidegger, *Contributions to Philosophy*, trans. Richard Rojcewicz and Daniela Vallega-Neu (Bloomington: Indiana University Press, 2012), 39-44.

make it more historically conscious. To be sure, much of Dewey's philosophy is geared expressly toward this end, and may even be considered poetic in many respects, but such thought is not instrumental.[26] Heidegger's "self-meditation" (which may be regarded as a kind of poetic thinking, though not necessarily a paradigmatic example) reflects on determination not in order to bring about any change in the physical environment, but so that the thinker may stand in a more knowing relation to her historicity, the inconspicuous (and for that reason worrisome) force that prods us along in our thought and actions.

One might object that in having such a purpose it is indeed instrumental to something, but if we allow this point then every kind of philosophizing—even the question concerning how many angels can fit on the head of a pin—must also be considered instrumental, in that it serves some human interest and purpose. All thought is certainly instrumental toward something, but not in such a way as to meet the criterion Dewey sets for it. It is precisely this criterion that I wish to call into question. If we demand of thought that it render physical modifications in the environment, we may be excluding equally valid forms of thought that move on more subtle levels.

Another objection that might be raised in response to this critique is the assertion that poetic thinking—and, for that matter, experience itself—is in fact physical, and thus that the changes it brings about must also be physical. I do not deny a certain version of this point, but let us not equivocate on the meaning of the word "physical." If by "physical" we mean embodied, then certainly poetic thinking and its effects are physical, since experience is necessarily embodied experience. But notice that, according to this definition, all forms of thinking (even the most groundless) qualify as generating physical alterations. Therefore, it would make no sense for Dewey to say that thinking ought to effect physical alterations in experience if he also thinks that it necessarily does so. Bearing this in mind, it is clear that Dewey does not use this word to signify anything like embodiment. Rather, by "physical" he seems to mean external, or independent of the agent, such that "physical alteration" consists in reconstructing the objective conditions of experience (recall the example of cleansing water).[27] Such alterations, however, are the domain of instrumental thinking; they are not of the kind that poetic thinking brings forth.

Up to this point, we have said very little regarding the positive characteristics of poetic thinking. It will be useful to highlight certain key aspects by illustrating in what ways poetic thinking differs from

[26] Cf., for example, John Dewey, *Reconstruction in Philosophy* (New York: Henry Hold and Company, 1920).

[27] The use of the words "external" and "independent" does not necessarily compromise the unity of the phenomenon of experience. External physical conditions are an inextricable part of experience, but we may nevertheless speak of them as being in a certain sense separate from the agent. It might be useful to make a distinction here between external physical conditions, which lie beyond the bodily limits of the agent (what I refer to as "physical conditions" in this essay), and internal physical conditions, the physical conditions of embodiment mentioned above. Both inhere in experience and are inextricably related, yet each has a distinctive phenomenological quality of its own.

its instrumental counterpart. We said earlier that Dewey conceives of experience as constituted by a complex of physical conditions, the same conditions that thought endeavors to modify. However, as Dewey himself acknowledges, experience is also constituted by significant conditions, that is, by conditions of meaning and value; it is saturated by the accrued (and ever accruing) significance of history and past experience. Rather than effecting any kind of physical alteration, poetic thinking sets in motion a significant alteration of experience. In much the same way as poetry and literature, it transforms the thinker's perception of the surrounding world, of things and others, altering the meaning of experience or bringing to light aspects of it that were previously hidden or overlooked. Such alteration of experience is most often unaccompanied by any corresponding physical alteration, but this does not make it any less actual or decisive. Poetic thinking certainly has effects, only they are not of the same kind as instrumental thinking. Whereas the latter intends its results, that is, projects the alterations it aims to implement in a clear-sighted manner, according to clear-set objectives, there is no equivalent conscious projection in the latter. In a certain sense, then, we may say that poetic thinking is purposeless, or even *useless* (according to everyday standards of utility). However, these should not be regarded as damning characteristics. I highlight them simply to illustrate that poetic thinking is unconcerned with technical expediency or manipulation of experience. Rather, it keeps within that domain which precedes all possible action and instrumentation—i.e., the interpretive domain. This is where its work is accomplished, and to demand that it bring about tangible, physical results is to foist upon it an incommensurable standard.

Conclusion

We find that Dewey's instrumental logic of thought gives us only a fragmentary picture of the potentiality of philosophical thinking. While he rightly recognizes that experience is constituted by significant as well as physical conditions, he demands of thought only that it generate physical alterations in experience, and as a result, significant alterations are regarded as mere by-products. This leads him to posit instrumentalism as accounting for the whole of philosophical thought. If, however, we are justified in claiming that thought may also aspire to significant alteration of experience as an end in itself, it becomes clear that instrumentalism is inadequate. In view of the aims of poetic thinking, physical alteration is of little concern and thought occupies itself with the interpretive domain, that is, with the significant conditions of experience. Thus, while instrumental thinking is legitimate when considered within its own limits, it fails to account for the whole of the possibilities of thought. Poetic thinking must be posited as its necessary supplement. ❖

Nietzsche's Society

Kristen Wells

Abstract: This essay asserts that Nietzsche proposes an important role for society within his ethics, and that this societal aspect has been greatly overlooked by Nietzsche scholars. By identifying a soul-state analogy and resemblance to virtue ethics, this essay contends that Nietzsche intends for societies and individuals to be seen as complementary parts of the will to power. Like Aristotle, Nietzsche prescribes an ideal society essential to greatness. By recognizing the importance of the role of society in Nietzsche's philosophy, Nietzsche scholarship is better positioned to consider new applications of his philosophical principles with his goals in mind.

A comprehensive approach to exploring Nietzsche's reoccurring consideration of society's value reveals that his ethics are meant to resemble a kind of virtue ethics and incorporate an Aristotelian soul-state analogy. The character and virtues Nietzsche identifies in individuals are inseparable and isomorphic to the societies' values, even though he often discusses the merits and faults of individuals and societies separately. A critical implication of this reading is that within Nietzsche's ethics, to strive for greatness, one must desire to create an ideal society. This supposition explains Nietzsche's preoccupation with evaluating societies and also suggests an aspiration of his philosophy: a better society. Thus, appreciating the full significance and role of society in Nietzsche's philosophy is necessary of Nietzsche scholarship if one is to correctly understand Nietzsche's values or carry on his philosophical tradition.

Throughout his works, Nietzsche is preoccupied with the question: what kinds of societies are valuable? Implicit in this question, Nietzsche makes two obvious claims. First, society has a role that is valuable. Second, there exists a method for determining value. In Nietzsche scholarship, explication of this second claim often overshadows the significance of the first.[1] Often, scholars approach the issue of society's value in Nietzsche's work by inferring from what they have already deconstructed; they first identify Nietzsche's ethics and then apply their own step-by-step methodology to determine the significance of the role of society. This approach results in shortsightedness. The persistence with which Nietzsche evaluates societies and the significance of this evaluation are lost unless the role of society is examined alongside his values.

The role of society has been overlooked due to several factors. The most overarching aspect is a failure to identify Nietzsche's soul-state analogy. Without realizing the connection between individuals' virtues and societies' virtues, Nietzsche's evaluation of society appears to be far less pervasive and one cannot see Nietzsche's objective change from assessing cultures to prescribing

[1] I have in mind here works by John Richardson, *Nietzsche's System* (New York: Oxford University Press, 1996) and Brian Leiter in *Nietzsche on Morality* (London: Routledge, 2002).

an ideal society. However, the role of the state in Nietzsche's middle and later period writings is easily missed. A surface reading of these works without consideration of his other works as a whole might suggest that Nietzsche is solely concerned with individuals. In fact, some passages from his notebooks seem to support the misconception that Nietzsche values individuals over society.[2] Moreover, wading through any of Nietzsche's deliberately provocative aphorisms posits difficulty. There are times when Nietzsche seems at best cryptic and at worst to contradict himself. This confusion is compounded further by Nietzsche's attempts to deter his readers from deconstructing his philosophy; he suggests that it would denigrate its essence. He does not want his project systematized, which is why he does not explain the soul-state analogy outright. However, he is writing to be understood by those who are able to appreciate the entire corpus of his work.

In his late notebooks, Nietzsche clearly questions the shape future societies should take. In this context, his favored term for society is culture (*Kultur*). One such passage states that "the Germans . . . have no culture yet" but are "becoming," that this is "a wish on which one can live, a matter of will," and that he and Germany "desire something more" from the German culture.[3] Writing either shortly before or after the first German unifaction, Nietzche is here describing a Germany that is essentially a new revolutionary society.[4] Nietzsche is excited to see a German culture that is growing, overcoming obstacles, and demanding more. Though much of Nietzsche's other published works express disappointment in German culture, his desire to prescribe a future ideal culture (in this case, for Germany) should not be dismissed as part of the mad ranting or fascist inlays that discredit the notebooks. It can also be found less explicitly stated in his earliest works. Beginning in his early writings, including *The Birth of Tragedy*, he aims to show how the culture of antiquity is superior to that of nineteenth-century Europe and summons Germany to embody the lost magnificence of Greek culture and thereby create a better society.[5]

Over time, Nietzsche's works convey his loss in confidence in modern society. However, he never loses admiration for antiquity, nor does he stop imagining a theoretical ideal society. Nietzsche's discussion of great societies evolves in his mature works from identifying examples of great societies to explaining the nature of great societies. He accomplishes this by imitating Greek culture, setting up his own soul-state analogy, which is

---

[2] Joe Ward, "Nietzsche's Value Conflict: Culture, Individual, Synthesis," *Journal of Nietzsche Studies* 41 (2011): endnote 3. Ward notes *KSA* 11:27[16], 12:5[108] on the value of the individual.

[3] Friedrich Nietzsche, *The Will to Power,* trans. W. Kaufmann and R. J. Hollingdale, (New York: Random House, 1968), 108.

[4] This passage could have been written between 1870 and 1890. The German unification officially occurred January 1871.

[5] Friedrich Nietzsche, *The Birth of Tragedy*, trans. Douglas Smith (New York: Oxford University Press, 2000).

first established in Plato's Republic.[6] In this analogy, Plato determines that a great society is a just society, and a state is just only if each of its parts—the money-making class, the helper, and the guardians—is engaged in that which corresponds to its nature, while working with others in mutual harmony.[7] According to Plato, man is by nature a political animal, while the state is the natural order where man is to be political. This constitutes what is known as Aristotle's political naturalism.[8] Unlike Plato, Nietzsche does not value the great society for political reasons. According to Nietzsche, "all great cultural epochs are epochs of political decline: that which is great in the cultural sense has been unpolitical, even anti-political."[9] Nietzsche is not interested in the government of the societies he evaluates. Rather, he is interested in their values, traditions, and how their structure or government is a part of the larger culture.

Nietzsche's soul-state analogy is not Plato's, but rather the interpretation by Aristotle who claims that the happiness and virtue of the state are the same as the happiness and virtue of the soul.[10] The difference between Plato's and Aristotle's analogies is that Aristotle's interpretation suggests the state and the soul share a nature that is not solely political but is the measure of every ethical decision. This seems to be true for Nietzsche, who considers some kinds of societies good because they reflect the highest state of being: "Society must not exist for society's sake but only as the foundation and scaffolding on which a choice type of being is able to raise itself to its higher task and to a higher state of *being*."[11] This quote suggests that societies are meant to benefit individuals, but the object of society is not only the individual but "being" itself, which includes more than the individual. Richard Schacht points out that Nietzsche also urges individuals to be "higher and freer," to "look beyond" all selfish considerations, and "pursue more distant purposes even under circumstances involving the suffering of others," as well as one's own suffering, for "through such sacrifice—meaning both ours and *our* neighbors'—we would strengthen and elevate the general sense of human power," even if we might "achieve nothing more."[12] According to Schacht, great individuals act truest to their nature when they contribute to their culture,

---

[6] Nietzsche appears to both idolize Socrates and blame him for the fall of Greek culture. See Daw-Nay Evans, "A Solution to The Problem of Socrates in Nietzsche's Thought: An Explanation of Nietzsche's Ambivalence Toward Socrates." (MA diss., Virginia Polytechnic Institute and State University, 2004).

[7] Plato, *Republic*, trans. G. M. A. Grube, ed. John M. Cooper (Indianapolis: Hackett, 1997), IV.

[8] In Aristotle, *Politics*, I, 2 (1253 18-29) the soul-state relationship is established in the opposite direction from the animal organism to the state.

[9] Friedrich Nietzsche, *The Anti-Christ, Ecce Homo, Twilight of the Idols, and Other Writings*, ed. Aaron Ridley and Judith Norman, trans. Judith Norman (Cambridge: Cambridge University Press, 2005), 4.

[10] In Aristotle, *Politics* VII, I, Aristotle explicitly agrees with the soul-state analogy in the very same terms as Plato states it in *Republic* IV but considers it in terms of virtue.

[11] Friedrich Nietzsche, *Beyond Good and Evil*, trans. W. Kaufmann (New York: Random House, 1966), 258.

[12] Friedrich Nietzsche, *Daybreak*, ed. Maudemarie Clark and Brian Leiter, trans. R.J. Hollingdale, (Cambridge: Cambridge University Press, 1997), 146.

which is the reason Nietzsche often talks about the higher-types as artists.[13] Therefore, individuals are agents whose highest state of being serves the creation of a high culture and vice versa.

The state and the individual have a shared nature (or purpose, depending on the metaphysical depth one reads into the will to power). One's thriving supports the other's thriving. They are isomorphic, which is the first premise of a soul-state analogy. Plato and Nietzsche also agree that those best suited to act as rulers are characterized by a similar internal structure.[14] Nietzsche's ultimate individual, the *Ubermensch*, is able to unify the opposing drives he bears just as the ideal society constructed in the soul-state analogy maintains stability amongst individuals of different classes.[15] Nietzsche clearly incorporates the structure of the soul-state analogy. Since he is concerned with culture, like Aristotle, he describes the nature of good in the structure of a virtue ethics: that which is good is a matter of character rather than abiding by set rules. Nietzsche constructs a kind of virtue ethics insomuch as he does not stipulate right actions to gain power; he only identifies power as virtuous.[16] At the same time, Nietzsche is certainly not the same kind of virtue ethicist as Plato or Aristotle. Bernd Magnus correctly rejects Walter Kaufman's reading of Nietzsche as an Aristotelian virtue ethicist. Aristotle's two main concepts, *eudaimonia* (happiness) and *phronesis* (practical wisdom) are completely unsuitable to Nietzsche.[17] Nietzsche clearly denies that happiness is the primary motivation of what is good.[18] He also reproves duty and logic.[19] The principle similarities between Aristotle's and Nietzsche's ethics that are important here are that they both strive for human greatness and measure an agent's quality in relation to how well it exemplifies its nature or essence.

The critical difference between Nietzsche's and Aristotle's soul-state analogies and ethics is what each believes to be the essence of life. Aristotle determines the human essence to be reason, whereas Nietzsche claims that it is the will to power. Nietzsche says:

---

[13] Richard Schacht, *Nietzsche* (London: Routledge & Kegan Paul, 1983), 469-73.

[14] In the *Republic,* the "rational part" of man must rule (441e); the philosopher alone is fit to rule the *polis* (487e).

[15] It is unclear if Nietzsche believes an *Ubermensch* could ever exist. It is possible that all people sometimes act highly or lowly, but only a few people are strong enough to value power.

[16] Michael Slote, "Nietzsche and Virtue Ethics," *International Studies in Philosophy* 30, no. 3 (1998): 23.

[17] Bernd Magnus, "Aristotle and Nietzsche: 'Megalopsychia' and 'Ubermensch'," in *The Greeks and the Good Life*, ed. David J. Depew (Fullerton: California State University, 1980), 262, and Walter Kaufmann, *Nietzsche: Philosopher, Psychologist, Antichrist* (New York: Vintage Books, 1968).

[18] Zarathustra asks, "What matters my happiness?" and recognizes that it is nothing which could "justify existence itself." Nietzsche, *Thus Spoke Zarathustra*, ed. A. Del Caro and R. Pippin, trans. A. Del Caro (Cambridge: Cambridge University Press, 2006), 3. Also see *The Anti-Christ*, 2.

[19] Christine Daigle, "Nietzsche: Virtue Ethics . . . Virtue Politics?" *Journal of Nietzsche Studies* 32 (2006): 4.

> [Anything which] is a living and not a dying body . . . will have to be an incarnate will to power, it will strive to grow, spread, seize, become predominant—not from any morality or immorality but because it is living and because life simply is will to power . . . 'Exploitation' . . . belongs to the essence of what lives, as a basic organic function; it is a consequence of the will to power, which is after all the will to life.[20]

The will to power is not only characteristic of individuals. It permeates every aspect of human life. Raymond Geuss believes its proper scope includes both smaller and larger entities that grow and become predominant through their own struggles and desires.[21] According to Geuss' argument, an individual is composed of many wills: He has a will, the company he works for has a will, his society has a will, etc. Societies are just as much a part of the will to power as individuals. Since accomplishments are relative to individuals' strengths, ethical classifications of character can only be described by the flourishing of individuals. This is why Nietzsche's late works define the will to power in relation to the individual. These individuals are only one dimension of the will to power. Their societies are another, and they affect the individuals' wills. Thus, the role of society is implicit in Nietzsche's discussions of individuals.

If the relationship between society's and individuals' will to power (Nietzsche's soul-state analogy) goes unrecognized, the role of society could seem important to Nietzsche only as an environment of the higher type. Some scholars, such as Brian Leiter, misread Nietzsche in this way, believing him to have abandoned the role of society to focus on the next virtuosos.[22] Leiter's reading refers to passages from Nietzsche such as, "a single human being can under certain circumstances justify the existence of whole millennia."[23] However, his conclusion is based on an incomplete analysis of this passage. Nietzsche certainly believes that some individuals are more valuable than their corrupt society, but this does not mean they are greater than the role of society. The same passage from *Twilight of the Idols* explains that great individuals, while often exceptions in their societies, are great because they act according to their nature and contribute to the creation of a high culture. In this same passage, Goethe is called a great individual because of his "grand attempt to overcome the eighteenth century through a return to nature, through a going-up to the naturalness of the Renaissance, a kind of self-overcoming on the part

[20] Nietzsche, *Beyond Good and Evil*, 259.
[21] Raymond Geuss, "Nietzsche and Genealogy," in *Morality, Culture, and History: Essays on German Philosophy* (Cambridge: Cambridge University Press, 1999), 12. Cf. Wolfgang Müller-Lauter, *Nietzsche: His Philosophy of Contradictions and the Contradictions of His Philosophy*, trans. David J. Parent (Chicago: University of Illinois Press, 1999), 181.
[22] Brian Leiter and Neil Sinhababu, *Nietzsche on Morality* (London: Routledge, 2002), 73–112.
[23] Nietzsche, *The Anti-Christ,* 49.

of the century."[24] Goethe, the great individual, overcame his society, but he did not secede from it. He inspired it to overcome itself through a return to nature.

Contra Leiter, Herman Siemens supports the position that society is equally as important to Nietzsche as great individuals. He points out that Nietzsche is most easily read as having concern for humanity as a whole. [25] Nietzsche claims that the corrupted morality of the lower-type is centered on individuals' well-being and happiness. He and other true philosophers ("we") reject this morality and value all life: "We . . . have opened our eyes and our conscience to the question where the plant 'man' has hitherto grown up most vigorously . . . everything evil, dreadful, tyrannical, beast of prey and serpent in man serves to enhance the species 'man'."[26] According to this passage, the future of humanity depends partly on suffering and destruction, which the lower-type believe are bad. Nietzsche is willing to sacrifice great individuals for the expansion of power.[27] Overall, power may increase through destruction, which is why happiness is not a primary component of Nietzsche's good.

Perhaps Leiter is mistaken about Nietzsche's disinterest in societies because he misinterprets Nietzsche's criticism of morality to mean he is an immoralist.[28] If Leiter is correct in this regard and Nietzsche has no objective measure of values, he need not create a soul-state analogy to express the ideal state of living. However, Nietzsche's criticism of morality does not suggest he is an immoralist. Nietzsche says the following about morality:

> In the main all those moral systems are distasteful to me which say: 'Do not do this! Renounce! Overcome thyself!' On the other hand I am favorable to those moral systems which stimulate me to do something, and to do it again from morning till evening, to dream of it at night, and think of nothing else but to do it *well*, as well as is possible for *me* alone! . . . I do not like any of the negative virtues whose very essence is negation and self-renunciation.[29]

Nietzsche approves of moral systems that motivate individuals toward growth. Even though this growth is relative to individuals, it is also the process of life perpetuating. It is the nature of life, the will to power, and the standard for morality. If Nietzsche is primarily concerned with individuals' accomplishments, as Leiter believes, he could not be an effective ethicist.

---

[24] Ibid.
[25] Herman Siemens, "Nietzsche's Critique of Democracy [1870–1886]," *Journal of Nietzsche Studies* 38 (2009): 30.
[26] Nietzsche, *Beyond Good and Evil*, 44.
[27] Friedrich Nietzsche, *The Gay Science*, ed. Bernard Williams, trans. Josefine Nauckhoff (Cambridge: Cambridge University Press, 2001), 349. Cf. Nietzsche, *Will to Power*, 688.
[28] Brian Leiter, "Nietzsche and the Morality Critics," *Ethics* 107, no. 22 (January 1997), 250–85.
[29] Nietzsche, *The Gay Science*, 304.

An individual's greatness is relative to his or her ability and circumstances. For instance, Napoleon's victories elevated his power and revolutionized his culture, but his power is not the same as power in general. By today's standards, his military tactics would accomplish little. Nietzsche's primary concern is the will to power of which societies are a part. The virtue ethics Nietzsche seems to construct is for the benefit of humans, but his standard of good encompasses more than humanity.

There are several levels at which Nietzsche believes a great society increases power. One controversial and well-known strain in Nietzsche's *Nachlass* characterizes the will to power as constituents of matter in the realm of physics.[30] Beyond this unsubstantiated notebook conjecture, several passages from the published works describe the human individual not with one agency but as having multiple drives, each with its own will to power.[31] Presumably, if there can be a synthesis of wills in the individual, there could also be further, more complex syntheses at the level of society.[32] John Richardson offers the foremost analysis of will to power as a synthesis of drives working together in stable tension.[33] Richardson, however, suspects that Nietzsche consciously avoids describing society this way, choosing to focus instead only on the exceptional individual.[34] He claims that "any society must be held together by values it can't see beyond so none can be that open-ended synthesis, always pressing to overcome itself, which is the Dionysian overman."[35] However, his interpretation of the society is too static. Societies cannot be so confined by their values as their values are always evolving, as shown by Nietzsche's need to write a genealogy of morality.

Bernard Reginster offers an intriguing insight into how the will to power benefits life, claiming that "the will to power, in the last analysis, is the will to the very *activity of overcoming resistance*."[36] Nietzsche's unpublished manuscript unmistakably claims that "all expansion, incorporation, [and] growth is striving against something that resists."[37] His published works describe expansion, incorporation, and growth as the will to power.[38] Hence, Reginster establishes his initial correlation between the will to power and resistance, and finds that his proposed relationship corresponds succinctly with Nietzsche's overall characterization of the good.

> What is good?—All that heightens the feeling of power, the will to power, power itself in man. What is bad—All that proceeds from weakness. What is happiness—The feeling that power increases—that resistance is overcome. Not

[30] Ward, endnote 3. Ward notes *KSA* 11:36[31], 12:9[98], 13:14[79], 13:14[95].
[31] Ibid., 7.
[32] This is similar to Siemens' conception of will to power.
[33] John Richardson, *Nietzsche's System* (New York: Oxford University Press, 1996), 44–52.
[34] Ibid., 51–2.
[35] Ibid., 141.
[36] Bernard Reginster, *The Affirmation of Life: Nietzsche on Overcoming Nihilism* (Cambridge: Harvard University Press, 2006), 127.
[37] Nietzsche, *Will to Power*, 704.
[38] Nietzsche, *The Gay Science,* 349 and Nietzsche, *Beyond Good and Evil,* 259.

> contentment, but more power; not peace at all, but war; not virtue but proficiency/virtue in the Renaissance style, virtue, virtue free of moral acid.[39]

In this passage, happiness is not only resistance, but resistance overcome. Additionally, will to power cannot only be the will to resistance against a perpetually frustrating obstacle. Otherwise, there could be no "expansion, incorporation, [or] growth." Nietzsche's conception of good is change and overcoming, which necessitates new and diverse resistance to overcome. The higher-type person is happy when she creates. That is her motivating desire, not hedonism or the feeling of happiness. The backbone of Nietzsche's values, the will to power, requires conflict and advancement.[40]

To understand Nietzsche's great society, one needs to re-evaluate the conventional portrait of utopia. For Nietzsche, a society is not great because it is peaceful or without need. A great society is an arena for meeting and overcoming resistance. Unsurprisingly, Nietzsche's outline of the great society resembles Plato's ideal society. Plato believes that few people achieve the ordering of the soul that results in highest virtue; the vast majority of people have misshapen souls.[41] Plato's conception of an ideal society concerns the flourishing of society, not its individuals. In such a society, not all citizens are great. It is an aristocracy in which lower types take care of menial tasks so that greater individuals can be involved in higher tasks.[42] Nietzsche claims that "caste-order . . . is necessary for the preservation of society, to make the higher and highest types possible,—unequal rights are the conditions for any rights at all."[43] In this way, lower types are essential to the great society but are not valuable like the higher types. Great individuals, like great societies, are creative and maintain a balance of power.

Though Nietzsche does describe his perfect society as having a similar structure to Plato's or Aristotle's societies, Nietzsche's society is not a product of temperance. According to Nietzsche's idea of the will to power, living a full life is not always accompanied by a satisfaction or a feeling of happiness. Nietzsche's ideal society is perfectly stable insomuch as it maximizes growth. People in this society aim to be innovators; they compete without there being a finish line. Each citizen is allowed to grow in whatever way he can manage: psychologically, physically, or mentally. Nietzsche believes that people do not all have the same strength or ability, which is why some people need and want to have subservient roles in society. The structure of this society prohibits these lower types from gaining enough influence to be able to destroy the society. Living well in the *polis* does not require practical wisdom; it requires an artist's imagination and the resolve to affirm the value of resistance.

While Nietzsche prescribes an ideal society, he only offers a theoretical structure of this society without committing himself to details. It is possible that

---

[39] Nietzsche, *Anti-Christ*, 2.
[40] Reginster, 126-47.
[41] Plato, *Republic,* 428d–e.
[42] Ward, 16.
[43] Nietzsche, *Anti-Christ*, 57.

the details of the perfect society are unrecognizable and unattainable, but in order to pursue greatness, we must strive for perfection nonetheless. To such an end, Nietzsche scholarship should be eager to acknowledge an exegetical analysis of the role of society as an attempt to do more with Nietzsche's philosophy. Granted, interpreting Nietzsche is a risky endeavor that has not always done him justice.[44] Still, we need to remember that Nietzsche does not want followers. He wants greatness.[45] Scholarship needs both an accurate understanding of Nietzsche's works and the ability to embrace the spirit of his work. To do either, one needs to acknowledge the unstated yet designated role of society in Nietzsche's ethics. ❖

[44] Walter A. Kaufmann, *Nietzsche: Philosopher, Psychologist, Antichrist* (Princeton: Princeton University Press, 1975).
[45] Nietzsche, *Ecce Homo*, IV.1.

Climate Ethics:

Individual vs. Collective Responsibility and the Problem of Corruption

Vesak Chi

Abstract: Anthropogenic climate change (ACC) has been described as a tragedy of the commons (T of C) by Baylor Johnson. Johnson argues that solutions to T of C scenarios reside in collective action rather than individual action, and that our moral obligation is to advocate for collective solutions to ACC. Marion Hourdequin argues that individual action can serve to promote collective action and in doing so it can also serve as an ethical obligation. I contend that individual action holds intrinsic value in lieu of its ability to counteract our susceptibility to the kind of moral corruption espoused by Stephen Gardiner.

The endeavor to provide substantive solutions to the crisis of anthropogenic climate change (ACC) is riddled with many and varying difficulties. At the very least there exist practical, logistical, theoretical, and philosophical obstacles that we as a people (both nationally and globally) must traverse before a real resolution to our collective predicament can be found. I choose now to focus on the philosophical hurdles, specifically the moral and ethical issues preventing us from achieving a solution to our climate troubles. Those who agree that a climate crisis exists and that action must be taken are faced with the problem of deciding exactly where their ethical obligations reside. I now understand, to some extent, the impact that greenhouse gas (GHG) emissions have on the environment. Should I not now restrict my own emissions so as to limit my personal contribution to ACC? Baylor Johnson argues that such views are fundamentally mistaken in his essay "Ethical Obligations in a Tragedy of the Commons." The problem of ACC, according to Johnson, is symptomatic of a tragedy of the commons, or a collective action problem. A tragedy of the commons (T of C) scenario is resolved not by "acting unilaterally," but by "acting as one of many in a cooperative scheme to address [the] problem."[1] I will argue that Johnson's view is potentially problematic because it makes a hierarchical distinction between unilateral actions and collective actions, and characterizes the former as inconsequential when done in isolation. With help derived from the work of Marion Hourdequin, I will argue that unilateral actions, even in isolation, are essential to solving the climate problem and should not be discounted.

I will begin by explaining Johnson's argument. As stated above, Johnson claims that the climate problem is essentially a T of C. The basic structure of a T of C is grounded in three central premises. Hourdequin provides an excellent

[1] Baylor Johnson, "Ethical Obligations in a Tragedy of the Commons," *Environmental Values* 12, no. 3 (2003): 272.

reiteration of Johnson's original explication of the premises, which accurately serves our purposes:

> 1. The only incentive players have is to maximise [their individual] benefits from use of the commons.
> 2. The only way players can communicate is by increasing or reducing use of the commons.
> 3. Use of the commons is shared, [however not all costs and benefits associated with use are shared.] Therefore:
>   a. *Costs (to the commons) of increased use* are shared, but *benefits from increased use* accrue to the individual . . .
>   b. *Benefits (to the commons) of reduced use* are shared, but *costs of reduced use* are borne by the individual . . .
>   c. Resources saved by one individual are available for use by any other user.[2]

Johnson states that "a T of C occurs when many *independent* agents derive benefits from a *subtractable* resource that is threatened by their *aggregate* use."[2] We can think of the atmosphere as a commons resource, the utilization of which consists of emitting GHGs for some benefit. There is a limit to the amount of aggregate GHG emissions the atmosphere can withstand before ACC begins to occur. Past this threshold point, we can say that the commons resource is being overused and the resulting ACC threatens the commons itself. It is important to note that individual GHG emissions result in no change in global temperature or in the occurrence of ACC. Rather, it is everyone's emissions combined that causes the harm resulting from ACC. When we combine the excessive use of the atmospheric commons, the absence of harm associated with individual actions, and the three premises outlined above, we can see that a T of C scenario obtains.

From here Johnson goes on to conclude that unilateral actions, which are individual actions not coordinated with some collective effort, are "ineffective in averting a T of C."[3] In our specific case, the unilateral actions are reductions in GHG emissions, and henceforth I will refer to these actions as unilateral reductions. Johnson's conclusion is grounded upon the idea that unilateral reductions without collective agreement will result in no alteration toward whether ACC will or will not occur. In order to prevent ACC, we must either prevent the T of C from obtaining or break out of it once it has obtained. Unilateral reductions lack the ability to do either, according to Johnson. In a T of C scenario, if an individual reduced his/her GHG emissions absent a collective agreement, then according to premise 2 and premise 3c above, that reduction would communicate to other individuals that additional resources are available for use. Consequently, according to premise 1, other individuals would be motivated to make use of the available resources for their personal benefit. Each individual who chooses to make use of the commons in such a way would be making the individually rational

[2] Johnson, 273.
[3] Ibid., 284.

choice, given the parameters of the T of C scenario. There are no assurances that any unilateral reductions will be mirrored by other unilateral reductions. Thus, any unilateral reduction will result in no change to the amount of GHG emissions made in aggregate, and so ACC remains inevitable.

Since unilateral reductions do not suffice as a means of averting ACC, the solution must reside in some other action. For Johnson, the solution is "to work for a collective agreement that could avert a potential T of C."[4] Once a collective agreement is established, it will bind everyone's actions and ensure that each individual will reduce their GHG emissions to sustainable levels or suffer possible repercussions and sanctions. Once this happens, the T of C is no longer in effect, and the commons will no longer suffer from overuse. Consequently, Johnson argues that advocacy for collective agreement is the primary ethical obligation for individuals if they seek to avert ACC. It is worth noting that Johnson seems to endorse a consequentialist moral theory that determines moral duties by reference to the success of their outcomes. In this sense, success is determined by avoidance of negative consequences, or promotion of overall utility.

Marion Hourdequin has advanced two arguments that challenge Johnson's claims of the limited ability of unilateral reductions. The first argument Hourdequin gives is the integrity argument, which aims mostly at invalidating Johnson's first premise.[5] According to Hourdequin, the principle of integrity can provide an alternate motivation for individual agents. Hourdequin initially calls the obligation for moral integrity "an obligation to avoid hypocrisy."[6] Instead of explaining the obligation with the negative connotation that hypocrisy implies, Hourdequin espouses a positive virtue of integrity that one should strive toward as an obligation. According to Hourdequin, integrity involves the idea of integrality, which is the internalizing of particular commitments which then become essential to the individual's identity. If a commitment is integral to an individual, then that commitment should be compatible or "*well integrated* with other commitments the individual holds."[7] So in the case of integrity with respect to addressing the climate problem, one must not only advocate for some collective agreement, but must "act also on a personal level to reduce her own emissions."[8]

Hourdequin's second argument is centered on an espousal of a Confucian interpretation of identity, which challenges Johnson's second premise. According to Hourdequin, "Confucian philosophy does not understand the individual as an isolated, rational actor," but rather as an entirely social being.[9] Thus, an individual's identity is a conglomeration of all

[4] Ibid., 283.
[5] Hourdequin also provides a brief explanation for how the integrity argument can counteract Johnson's second premise, but I choose to disregard that facet of the integrity argument. The reason I do so is because Hourdequin's Confucian argument does a more thorough job of invalidating Johnson's second premise, while her integrity argument neatly challenges the first.
[6] Hourdequin, 448.
[7] Ibid.
[8] Ibid., 449.
[9] Ibid., 452.

the social interactions and relationships that individual holds. If an individual can be interpreted in this manner, then any or all unilateral actions made by an individual can influence others within a shared social contact. Further, such an individual learns about moral and ethical actions and behaviors through observation and interaction with surrounding people. In this sense, the Confucian interpretation of the self can effectively nullify the restrictions in communication inherent in the T of C framework.

In response to these arguments, Johnson, in a later work, altered some of his views on the importance of unilateral actions. Most importantly, Johnson now agrees that "unilateral reductions can be valuable" insofar as they complement and support the call for collective action and agreement.[10] This is an important reevaluation, because it concedes the fact that unilateral actions have a communicative property such that they can help to influence the behavior or views of others. Though Johnson does make this concession, he states that we must clearly distinguish "unilateral reductions *in isolation* from unilateral reductions *in combination* with a richer strategy for communication."[11] In essence, Johnson says that the value in unilateral reductions resides in their ability to combine with and promote the social advocacy for collective agreement. But, when unilateral reductions are done in isolation and communicate no morally salient ideas to others, then they continue to be morally neutral or inconsequential. Finally, Johnson also makes a hierarchical distinction between the two forms of action. Advocacy of collective action, insofar as it is the primary means of effecting change, continues to hold precedence over any communicative unilateral reductions.

Hourdequin issues a reply to Johnson that primarily argues against the hierarchical distinction Johnson makes between unilateral reductions and the advocacy of collective action. She argues for the increased importance that must be placed on unilateral reductions. Hourdequin emphasizes that "*individual emissions reductions can themselves contribute to the generation and stabilization of effective collective schemes*" and as such, the distinction between the two is not at all decisive.[12] In response to the distinction between unilateral reductions done in isolation and those made with the intent to complement social advocacy, she says that "barring an unusual degree of isolation from others, very few 'unilateral reductions' will be truly private."[13] In essence, her argument against this distinction is simply to say that most acts of unilateral reduction are not entirely isolated. However, it is important to note that Hourdequin does not provide a clear and salient argument for the importance of unilateral reductions even in complete isolation. Even though she may not espouse any clear argument in her response to Johnson, perhaps we may look back to her argument on integrity in order to facilitate the creation of such an argument. For instance, she could perhaps pin the importance of

[10] Baylor Johnson, "The Possibility of a Joint Communiqué: My Response to Hourdequin," *Environmental Values* 20, no. 2 (2011): 150.
[11] Ibid., 150.
[12] Marion Hourdequin, "Climate Change and Individual Responsibility: A Reply to Johnson," *Environmental Values* 20, no. 2 (2011): 162.
[13] Ibid., 159.

unilateral reductions in isolation upon the moral importance of maintaining integrity and avoiding hypocrisy.

If we are to show the importance of unilateral reduction in isolation, our argument cannot depend on challenging Johnson's second premise, that of communication. Rather, integrity must be something that is morally significant in its own right regardless of its influence on others. This is exactly what Hourdequin does when she describes the positive connotations of the obligation of moral integrity. If I am truly committed to addressing the climate problem, I must make my commitment integral to who I am as a person, and my commitment must be integrated into all of my activities so as not to create conflict among my actions. So, if I value integrity, then I must value the personal commitment of unilaterally reducing my emissions even in isolation. This seems rather straightforward. But there is a problem: what Hourdequin merely provides is a contrary principle to challenge Johnson's principle of self-interest. Her argument is to pose an alternative principle and espouse its qualities and hope that in doing so it will prove to be more of an incentive than self-interest and personal utility. This form of argumentation does not establish in a compelling fashion why we should adopt integrity over self-interest. But I believe there is a different road to be taken.

Hourdequin originally classified the obligation of moral integrity as an obligation to avoid hypocrisy. She chose not to pursue this obligation because of its negative connotation. However, I believe that a very strong argument resides down this path. When we view someone as being hypocritical, we often make moral judgments about their hypocrisy. When we judge a person as being hypocritical, we are saying that the person has made some mistake or contradiction in their behavior, and as such they are subject to moral reprobation. When a person advocates for collective action to solve our climate problems by restricting everyone's GHG emissions to sustainable levels, and while doing so continues to emit in a wanton fashion, we judge this person negatively. Yet the problem here is much more disconcerting than a mere negative moral judging.

Stephen Gardiner in his work "A Perfect Moral Storm: Climate Change, Intergenerational Ethics, and the Problem of Corruption" argued for what he referred to as a "distinct problem for ethical action on climate change."[14] This problem was that of moral corruption. Gardiner believed that ACC posed such a unique problem that there were many facets of difficulties that had to be solved before a solution could be found. The culmination of all of these factors led Gardiner to describe "Climate change [as] a perfect moral storm."[15] The meaning behind this categorization is that the large problems surrounding ACC "exacerbate and obscure a lurking problem of moral corruption."[16] This moral corruption makes us susceptible to distraction, complacency, self-deception, selective-attentiveness, and hypocrisy. All these

[14] Stephen Gardiner, "A Perfect Moral Storm: Climate Change, Intergenerational Ethics and the Problem of Moral Corruption," in *Climate Ethics*, ed. Stephen M. Gardiner, Simon Caney, Dale Jamieson, and Henry Shue (New York: Oxford University Press, 2010), 94.
[15] Ibid., 88.
[16] Ibid.

things work towards preventing action and resolution to the climate change problem. If we categorize the hypocrisy inherent in Johnson's argument as being linked to moral corruption then we can begin to espouse an argument against him that is perhaps stronger than integrity alone. If Gardiner is right and moral corruption of this sort inhibits successful solutions to the climate crisis, then holding such a hypocritical view will not easily lead to successful implementation of collective action. We cannot truly expect to efficiently and successfully advocate for collective action when we ourselves are hindered by moral corruption such that we do not incorporate the actions we advocate into our behavior. It does not matter whether the hypocrisy is a result of moral corruption or not, it still results in the propagation or continuation of moral corruption. If this moral corruption prevents or even inhibits the successful advocacy of collective action, then Johnson's consequentialist argument falls apart.

As I have mentioned above, Johnson's argument hinges upon a consequentialist framework wherein moral obligations coincide with actions that serve to promote overall utility. Johnson believes that collective action is hierarchically superior to unilateral reductions because collective action serves to avert ACC by preventing or defeating a T of C scenario, whereas unilateral reductions lack this capability. However, when we take into account the moral corruption espoused by Gardiner, we can see that any collective action that is bereft of unilateral reductions is insufficient in regards to effecting a successful solution. If this is so, then it would seem that any hierarchical distinction between collective actions and unilateral reductions is patently mistaken. For it would seem that both collective actions and unilateral reductions are necessary for the successful resolution to a T of C. If this is the case, then Johnson must either relent to the equal importance of unilateral reductions or abandon his consequentialist framework.

In summation, there is significant moral value in an act of unilateral reduction in isolation. This value lies in solidifying and unifying our moral obligations, both collective and unilateral, to address ACC. In doing so, we will have taken the initial steps necessary in order to nullify the looming threat of moral corruption. It is my hope that taking these steps will eventually result in a more efficient and pervasive collective agreement that is uninhibited by our invariable susceptibility to moral corruption. A strong and equal emphasis must be placed on both collective action and unilateral reduction in order to defeat this T of C scenario and resolve our climate troubles. Any hierarchy placing one set of actions above the other will insufficiently preclude the possibility of moral corruption and only hinder our efforts for a climate resolution. It is not enough to promote unilateral reductions based upon their communicative properties. There is value in the integrity established by unilateral reductions regardless of isolation. Nor is it enough to promote only collective action while failing to ingrain within ourselves the beliefs inherent in such advocacy. If we are to truly seek a path leading to the resolution of our current climate dilemma, we must proceed forward as individuals and as a collective with the full commitment to strive toward a future unhindered by the shadow of

moral corruption and the contingent repercussions of anthropogenic climate change. ❖

Carruthers and Constitutive Self-Knowledge

John C. Hill

Abstract: In his recent book, *The Opacity of Mind*, Peter Carruthers advances a skeptical theory of self-knowledge, integrating results from experimental psychology and cognitive science.[1] In this essay, I want to suggest that the situation is not quite as dire as Carruthers makes it out to be. I respond to Carruthers by advancing a constitutive theory of self-knowledge. I argue that self-knowledge, so understood, is not only compatible with the empirical research that Carruthers utilizes, but also helps to make sense of these results.

Introduction

In his recent book, *The Opacity of Mind*, Peter Carruthers advances a skeptical theory of self-knowledge, integrating results from experimental psychology and cognitive science. Carruthers argues that our knowledge of ourselves is always mediated by a mindreading faculty that never directly perceives our attitudes. Indeed, recent work in experimental psychology and cognitive science challenges philosophical conceptions of self-knowledge through an emphasis on ways in which the sub-personal processes beyond our conscious awareness go awry.

In this essay, I want to suggest that the situation is not quite as dire as Carruthers suggests. Self-knowledge is tied to normative notions like commitment and agency in ways that other-knowledge is not. As a result, a more thorough understanding of ourselves, warts and all, can help us see where we make mistakes and how to improve on them. In this essay, I proceed as follows: first, I motivate a constitutivist theory of self-knowledge.[2] Next, I introduce and evaluate Carruthers proposed theory of self-knowledge, the interpretative sensory-access theory, and explore its skeptical conclusions. I

[1] Peter Carruthers, *The Opacity of the Mind* (Oxford: Oxford University Press, 2011).

[2] I adopt the term 'constitutivism' with some hesitation. In the literature on self-knowledge, constitutivism sometimes refers to a view of self-knowledge according to which it is constitutive of belief (or other propositional attitudes) that if one believes that p, then one believes that one believes that p. In other words, the nature of belief itself entails self-knowledge of beliefs. For examples of this view, see Sydney Shoemaker, "Self-Knowledge and 'Inner Sense'," *Philosophy and Phenomenological Research* 54 (1994): 249–314; Eric Schwitzgebel, "Knowing Your Own Beliefs," *Canadian Journal of Philosophy* 35 (2011): 41-62; and Declan Smithies, "A Simple Theory of Introspection," in *Introspection and Consciousness*, eds. Declan Smithies and Daniel Stoljar (Oxford: Oxford University Press, 2012), 259-93. Another candidate term, for instance, 'contructivism,' has anti-realist connotations that I would like avoid (see below). Consequently, I have settled on the term 'constitutivism', but keep in mind that the type of constitution at issue is self-constitution, rather than the constitution of belief itself.

argue that constitutivism, because it reconceives the nature of self-knowledge, is especially suited to incorporate such findings while preserving an important place for our ability to determine the state of our minds.

Constitutive Self-Knowledge

The constitutive theory of self-knowledge is a distinctive theory of the asymmetries that hold between one's relation to one's own mind and one's relation to the minds of others. The core claim of the constitutive account of self-knowledge is that knowledge of one's state of mind is a process of self-constitution, rather than self-discovery.

The constitutive account of self-knowledge can be described in opposition to an observational account of self-knowledge. According to an observational account of self-knowledge, when one ascribes an attitude to oneself, one reports an already given mental state which one observes via the mental faculty of introspection. Accordingly, observational accounts of self-knowledge explain one's first-person authority vis-à-vis one's own attitudes in terms of a privileged perspective from which one observes those attitudes. However, one's ascriptions of mental attitudes to others lack such special access, and consequently are less epistemically secure.

As many philosophers have argued, such observational theories of self-knowledge cannot account for the immediacy of our self-knowledge.[3] That is to say, when one ascribes an attitude to oneself, one typically does not attempt to gather evidence or try to observe what one thinks; instead, one merely ascribes that one is in such-and-such a mental state. If our practice of self-ascription is not wholly unwarranted, then the authority of our self-ascriptions must derive from elsewhere.

In contrast to observational accounts of self-knowledge, constitutive accounts of self-knowledge argue that one's first-person authority results from one's capacity to effectively constitute one's own attitudes. As Finkelstein puts it, "typically, what I say or think about my own mental state plays a constitutive role in determining what it is."[4] In other words, one has authority over one's mental life because one can make up his or her mind about some state of affairs. For instance, Susan knows that she believes her friend Amy is trustworthy, not because she has some special evidence which bears on her mental state, but because in considering the question of whether she believes Amy is trustworthy, she forms a novel attitude toward Amy's trustworthiness. In effect, she makes up her mind regarding Amy's trustworthiness.

According to the constitutive view, self-knowledge is the result of our capacity to commit ourselves to various attitudes. For instance, Bilgrami argues

---

[3] The constitutivism I endorse is inspired in part by Richard Moran, *Authority and Estrangement: An Essay on Self-Knowledge* (Princeton: Princeton University Press, 2001). Cf. Tyler Burge, "Individual and Self-Knowledge," *Journal of Philosophy* 85, no. 11 (1988), 649-63 and Akeel Bilgrami, *Self-Knowledge and Resentment* (Cambridge: Harvard University Press, 2006).

[4] David Finkelstein, *Expression and the Inner* (Cambridge: Harvard University Press, 2003), 28.

that "to desire something, to believe something, is to think that one ought to do or think various things."[5] When one ascribes an attitude to oneself or makes an avowal concerning one's state of mind, one undertakes a commitment to reason and deliberates in accordance with the normative requirements associated with adopting such an attitude. For instance, to constitute oneself as someone who desires social justice is to commit oneself to bringing about social justice. It is the capacity to commit oneself that explains the asymmetry between self-ascriptions and other-ascriptions. Importantly, one is responsible for one's commitments. In the constitutive view, one's authority is a result of one's responsibility for his or her state of mind.[6]

At this point, one might worry that the constitutivist theory entails that whatever one thinks about one's state of mind completely determines its character. Interpreted in this way, the constitutivist theory of self-knowledge is highly unrealistic. Finkelstein, for instance, complains, "constitutivism has the effect of misrepresenting the subject's responsibility for mental states about which he speaks with first-person authority . . . If my remarking that I had a headache made it the case that I did have a headache, then sympathy would not be appropriate in response to such a remark."[7] Or, consider Coliva, who writes regarding constitutivism, "psychological self-ascriptions such as 'I believe/desire/intend/wish/hope that P' do bring into existence the corresponding first-order mental states . . . On this model, there would be a sense in which it is literally true that we make up or create our minds."[8] In other words, constitutivism maintains that we have control over our mental lives that we seemingly do not. What is more, it seems as though constitutivism ends up committed to a stronger form of infallibility than observationalism. Considering that constitutivism is compatible with substantial fallibility, this is a poor result.

However, the process of self-interpretation that is involved in self-constitution is constrained. The act of self-interpretation is constrained by one's antecedent attitudes. Moran accurately locates the perceived problem in the language used to describe the process of self-interpretation:

> Verbs such as 'describe', 'interpret', and the like are fated to equivocate between a use that expresses one's genuine sense of how things are, with the same kind of commitment as belief, and a different, noncommittal use denoting an ordinary activity. Favoring verbs in this latter

[5] Bilgrami, *Self-Knowledge and Resentment*, 213.

[6] Cf Keith Frankish, *Mind and Supermind* (Cambridge: Cambridge University Press, 2004); Keith Frankish, "Dual Systems and Dual Attitudes," *Mind and Society* 11 (2012), 45-51; Peter Carruthers, *The Opacity of the Mind* (Oxford: Oxford University Press, 2011); and Peter Carruthers, "On Knowing Your Own Beliefs," in *New Essays on Belief: Structure, Constitution and Content*, ed Nikolaj Nottelmann (New York: Palgrave MacMillan, forthcoming).

[7] Finkelstein, 52.

[8] Annalisa Coliva, "One Variety of Self-Knowledge: Constitutivism or Constructivism," in *The Self and Self-Knowledge*, ed. Annalisa Coliva (Oxford: Oxford University Press, 20

> sense serves to dramatize the idea of self-transforming redescription, and to obscure its genuine basis.[9]

The upshot is this: there are two different senses of 'interpret' with different consequences for the process of self-constitution. The first sense allows for a great degree of freedom in process, by analogizing self-interpretation with an act of describing, unconstrained by cognitive considerations. The second sense constrains the process of self-constitution by connecting it with commitment. In self-interpreting, an individual forms a belief about himself or herself, sensitive to the same epistemic considerations as any other belief. For example, an individual who comes to see a cigarette, which was once formerly craved, as a harmful, addictive substance, adopts a new belief about its desirability and so creates a desire to avoid smoking. Such an individual does not merely describe smoking as unhealthy or undesirable, but forms a cognitive commitment not to smoke instead.

Carruthers's Skeptical Theory: Interpretative Sensory Access

Carruthers argues that the system by which one ascribes attitudes to oneself is the same system by which one ascribes attitudes to others. Following the extensive literature on so-called "theory of mind," Carruthers calls this system the mindreading faculty. Carruthers dubs his account of self-knowledge the "interpretative sensory-access theory" or ISA, because he contends that this system only has access to sensory or perceptual information about one's own bodily movements, one's physiological arousal, and one's inner speech. Carruthers' ISA theory has skeptical conclusions because it suggests that our access to our thoughts is always mediated by the mindreading faculty, which lacks direct access to our attitudes.

Allow me to describe the ISA theory in a bit more detail. Carruthers develops his theory in light of well-established empirical theories in the foundations of cognitive science. In particular, he assumes a global broadcast architecture of the human mind, proposed initially by Baars.[10] This means that Carruthers assumes the mind is differentiated into two classes of cognitive systems. There are sensory systems that receive sensory information and broadcast it to an array of consumer systems, which, according to Carruthers, include the mindreading faculty. In Carruthers' words, "human mental life consists of islands of conscious events surrounded by seas of unconscious processing."[11] As a result, there is no special epistemic relation grounding our first-person authority in relation to our own minds.

Carruthers claims his theory explains data gathered by experimental psychologists and neuroscientists on confabulation, or the process of filling in gaps in one's knowledge with false or unjustified judgments. Indeed,

---

[9] Moran, 53.

[10] Bernard Baars, *A Cognitive Theory of Consciousness* (Cambridge: Cambridge University Press, 1988).

[11] Carruthers, *The Opacity of Mind*, 50.

confabulation is a direct obstacle to self-knowledge as traditionally conceived, for it suggests that we lack first-person authority in relation to our states of mind. As Carruthers himself puts it, "[s]ince the [ISA] theory claims that our only access to our thoughts and thought processes is interpretative, relying on sensory, situational, and behavioral cues, there should be frequent instances where presence of misleading data of these sorts leads us to attribute attitudes to ourselves mistakenly."[12]

There is widespread literature on confabulation. Allow me to briefly present a particularly vivid example. Brasil-Neto et al. induced motor movement through direct stimulation of the motor cortex, bypassing decision-making systems of their subjects.[13] They made their subjects wear headphones, and then instructed their subjects to raise their index finger when they heard a clicking noise. However, they sometimes induced this finger movement artificially in their subjects without the presence of audio. Their subjects self-identified these finger movements as intentional decisions to move their fingers. The result of these considerations is that we are prone to misinterpret our actions and misidentify the reasons for which we act. Carruthers takes these examples to entail a strong form of skepticism about our access to our own minds. Indeed, they seem to undermine a constitutivist account of self-knowledge. But below, I will argue that if properly understood, these results can be incorporated into a constitutivist account of self-knowledge.

Responses to Carruthers

The constitutivist agrees with Carruthers that one's self-knowledge of one's state of mind is interpretive. Where Carruthers and the constitutivist disagree is in the implications of interpretation of first-person authority. Carruthers takes the interpretative nature of self-knowledge to undermine first-person authority, while the constitutivist takes the interpretative nature of self-knowledge to be essential for first-person authority.

The constitutivist also agrees with Carruthers that we can be influenced by mental states that work beneath the surface, but the constitutivist does not draw the same skeptical conclusions from this fact. It is important to distinguish mental states of which we have knowledge from mental states of which we are consciously aware.[14] I argue that we have knowledge of certain mental states without necessarily being consciously aware of them, and that this consciousness of our minds constitutes an important type of self-knowledge.

The distinction between knowledge and conscious awareness of states of mind can be articulated by means of an example. First, consider Cam, who is consciously angry at his brother, Matt. Cam experiences the phenomenal heat of anger; he feels a bodily reaction when his brother enters the room, and he is aware of considerations that justify his emotion toward his brother

[12] Ibid., 325.
[13] Joaquim Brasil-Neto et al., "Focal Transcranial Magnetic Stimulation and Response Bias in a Forced Choice Task," *Journal of Neurology, Neurosurgery, and Psychiatry* 55 (1992): 964-6.
[14] Cf. Finkelstein, *Expression and the Inner*, 20.

(for instance, his brother has done something to offend him). Second, consider Ned, who is not consciously angry at his brother. Instead, he may simply notice a tendency to think negative thoughts when his brother is around. Or perhaps their mutual friend tells Ned that his body language becomes aggressive and tense when he describes a meeting with his brother.

Ned considers his stance toward his brother and decides that he is angry at him. Although Ned knows of his anger toward his brother, he does not consciously experience it as anger. Instead, he might interpret his anger as a mere mood or inclination based on external facts about his current situation, for instance, the fact that he skipped breakfast. In other words, he lacks conscious self-knowledge of his state of mind. Nevertheless, his knowledge of his anger, once it is brought to light, opens the door to alternative ways of understanding it. For instance, Ned might come to recognize his behavior toward his brother as manifesting a non-conscious anger toward him, and consequently come to see his actions toward his brother as unwarranted.

Such a realization may not immediately change Ned's dispositions to act in relation to his brother (e.g., he might feel motivated to criticize his brother upon hearing his suggestions), but mere recognition of this aspect of his state of mind, though not directly accessible in some infallible, self-presenting sense, is still a crucial type of self-knowledge for Ned's process of self-constitution. Without consciousness of his anger, he would lack the reflective distance to commit himself to alternative ways of being.

It is not open to Carruthers to deny the existence of knowledge of our mental states in the sense described above. Indeed, Carruthers' own theory is a testament to how knowledge of the workings of our mind is possible, even if we cannot have conscious awareness of these workings. However, Carruthers might argue that the type of knowledge I am describing is unimportant. According to this line of thought, only consciously experienced mental states play a role in decision making.

Such an objection misses the importance of knowledge of our minds and their workings, along with the normative elements of commitment, in our practical reasoning. Speaking on the compatibility of observational and constitutive accounts of self-knowledge, Moran captures these considerations quite nicely:

> Neither perspective denies the truths of the other. The assertion from the Deliberative stance that "I am not bound by my empirical history" is not in any way a denial that the facts of my history are what they are. It does not deny either the truth of these claims or their relevance to the question at hand; but it does deny their completeness and, in a word, their decisiveness.[15]

In other words, Moran argues that knowledge of the empirical reality of our condition, of the ways in which we are prone to err when thinking about ourselves, is essential for coming to a reasoned stance about how we ought

[15] Moran, 163.

to proceed in our lives. Nevertheless, such empirical knowledge cannot fully determine how we ought to proceed. As Moran describes, it is incomplete. We have a responsibility to constitute ourselves, not simply to follow through on some projected trajectory. We are presented with an external world that requires us to make choices and perform actions. The option to simply submit to our habitual dispositions is not really an option. To put it another way, self-knowledge is deliberatively indispensable.

Note that constitutivism about self-knowledge, so conceived, neither requires transparency in the Cartesian sense advocated by Carruthers, nor does it deny interpretation. What is more, it acknowledges that self-knowledge is a genuine cognitive achievement. In other words, self-knowledge must be earned through constant revaluation of one's empirical conditions and one's commitments. But, as Moran points out, a more thorough knowledge of our circumstances need not undermine the core of the constitutivist position.

Conclusion

I have argued that Carruthers's skeptical conclusions about self-knowledge are unwarranted. The constitutivist account of self-knowledge for which I have argued recognizes that our knowledge of ourselves can be difficult to acquire, but it need not be any more inherently problematic than any other kind of knowledge. The essence of constitutivism is the thesis that self-knowledge is a process. Taking this view seriously allows us to appreciate not only the shortcomings, but also—and more importantly—the virtues of self-knowledge and its significance in our lives. ❖

Not So Innocent:

An *Akratic* Reading of Leibnizian "Judgment"

Oda Storbråten Davanger

Abstract: Leibniz seeks to establish the tenability of faith and reason in his moral philosophy through a tripod of thought, consisting of 1) fundamental human goodness; 2) human error in judgment; and 3) that God is just. A difficulty arises concerning how God can justly punish human beings if they always will what is Good. By considering *akrasia*, which occurs when error is committed despite its clear nonconformity with the Good, and examining the Leibnizian concept of "judgment," Leibniz's tripod can be upheld.

When God makes a choice, it is through his knowledge of the best; when man does so, he will choose the alternative that seems to be best.[1]

—G. W. Leibniz, 1707 letter to Coste

Introduction

To give an account of a perceived difficulty in Leibniz's moral philosophy, I have identified a tripod of thought. The tripod consists of three pillars that hold up Leibniz's ethics, which may topple if one of these pillars is shown to be untenable with the others. The pillars of the tripod, in simple terms, consist of 1) Leibniz's belief in the fundamental goodness of humanity, or rather, his belief that humans always will what is Good; 2) the notion that human beings make judgmental errors because of their limited knowledge, which may lead people to unknowingly commit error;[2] and finally, 3) the idea that God is Good and therefore just in his administration of punishment for error. For Leibniz, the use of reason and knowledge combined with faith in the Good God leads to the practice of morality—an ethical navigation between the known and unknown.

Sometimes what we judge to be the best choice based on our limited knowledge of things does not, in reality, conform to the Good. The issue at stake is how Leibniz can hold human beings morally accountable to God's just punishments if he simultaneously posits that human beings always will what is Good and mistakenly err. Because the human will always wills what is Good, it cannot be accountable for error it did not intend. This difficulty becomes apparent as we consider how unjust it would be for God to punish people for their wrongdoings if they thought they were doing something Good. As such, Leibniz's position only succeeds if the one who is punished knows why, namely, that he or she is guilty of committing error.

In the first section, I give an account of what I have identified as Leibniz's moral tripod and explain how its philosophical tenability relies on

[1] John Hostler, *Leibniz's Moral Philosophy* (London: Duckworth, 1975), 31.

[2] Leibniz often refers to errors as "evil."

accountability. Next, I find that there is a divide between Leibniz scholars regarding their understanding of how human beings commit error, whether error is due to the presence of mistakes or passions in judgment. Neither tradition, however, explores the relation between *akrasia*, viz., committing error knowingly, and accountability in Leibniz's concept of judgment. Third, I give an account of Leibniz's struggle with the tenability of reason and faith, and that, while the will is bound to what is Good, judgment may have the capacity to steer volitions in accord with the passions. I conclude that more research should be done to incorporate Leibniz's psychological elements into how scholars understand his moral philosophy, which would provide for a more tenable tripod.

The Tripod

In order to maintain some of his fundamental presumptions—that God exists and is Good, and that reason is tenable—Leibniz must somehow acknowledge that guilt is inherent to humankind. Frankel identifies Leibniz's tripod as an attempt to preserve "human freedom, divine freedom, and contingency," and seeks to combine faith and reason:

> He must preserve *human freedom* firstly because 'it is in freedom that we seek the reason for praise and blame', and secondly, so that we, rather than God, may be held morally responsible for our sins. He must preserve *divine freedom* in order to preserve God's moral perfection, which requires it to choose freely to do what it knows . . . to be the best.[3]

In other words, free acts must be subject to the agent's control rather than being accidental or constrained.[4] In order to protect the validity of God's moral perfection and the justness of His punishments (and rewards), human beings must be accountable for error. As such, reason is at stake as it applies to the freedom of humanity while divinity is also at stake insofar as the justness, namely, the Goodness of God is questioned. According to the first pillar of the Leibnizian tripod, the human will is a divine inheritance and thus humans, like God, always will what is Good.[5] Human beings have inherited the ability to use knowledge and reason to direct their wills toward the Good, but have inherited these perfections only limitedly.[6] The more they act in accordance with knowledge and "right reason," [7] the more

[3] Lois Frankel, "Being Able to do Otherwise: Leibniz on Freedome and Contingency," in *Gottfried Wilhelm Leibniz: Critical Assessments*, ed. R. S. Woolhouse (New York: Routledge, 1994), 284.
[4] Ibid., 285.
[5] Bertrand Russell, *A Critical Exposition of the Philosophy of Leibniz* (London: Alden Press, 1975), 191.
[6] G. W. Leibniz, "Discourse on Metaphysics," in *Discourse on Metaphysics and Other Essays*, ed. Daniel Garber and Roger Ariew (Indianapolis: Hackett, 1991), §1.
[7] G. W. Leibniz, "Letter to Magnus Wedderkopf, May 1671," in *Philosophical Papers and Letters: A Selection*, ed. and trans. Leroy Loemker (Chicago: Chicago UP, 1956), 227.

free are human beings, for it is then that they allow their wills to contribute to the manifestation of what is Good.[8]

A person who knows through reason what is Good but chooses to act according to apparent goods and passions rather than the actual Good is *akratic. Akrasia* is a term used to describe a situation where "an agent fails to adopt in practice what he sincerely judges to be the best course of action."[9] For example, someone may choose to smoke cigarettes while pregnant, despite knowing that this act is wrong, viz. an error. In contrast to the Aristotelian understanding of *akrasia*—that the *akratic* person is overcome by desire—Leibniz believes that the *akratic* person actively chooses the less good action "in accord with reasons, albeit ones that would not generally be regarded as valid."[10] The smoking pregnant woman, for example, may reason that her smoking habits are beneficial to the pregnancy because quitting would create horrible levels of stress, although she very well knows that this reason is not valid. The question pertaining to the Leibnizian tripod is how to account for *akrasia* if human beings always will what is Good. Examining the Leibnizian concept of judgment can illuminate how *akrasia* and human goodness can coexist without contradiction.

According to Leibniz, judgment plays a special role in volition. For Leibniz, volition is distinct from the will. The two elements of Leibnizian volition are *conatus*, translated as "striving, essence, or desire," and *opinio,* namely judgment.[11] In other words, 1) human beings will the Good, and 2) in order to conclude in a judgment they must deliberate by reasoning about what the Good is in any given situation. Leibniz held that the human will cannot alone initiate volition, which also relies on judgment.[12] For Leibniz, *conatus* is a striving that follows *opinio*, which is responsible for identifying the Good.[13] As such, volition is the product of the relation between willing and judging. How one understands the concept of judgment is important to one's reading of the second pillar of the Leibnizian tripod, because it influences accountability and dictates whether or not error is strictly committed unknowingly.

For Leibniz's tripod to be tenable there must be some accountability that renders human beings responsible for their actions. I claim that such accountability is located not in the will, but in the thinking of volition, namely judgment. Leibniz's concept of judgment may preserve human goodness while it also accounts for the error that is derived from the misuse of reason. Perhaps this misuse, or rather, the "improper use of ideas [that] gives rise to several

[8] Leibniz is drawing from Augustine's "The Free Choice of the Will," which posits that freedom is not achieved by being free of a master, but in following the right master. The right master is reason and knowledge, not passion. *The Fathers of the Church*, vol. 59, trans. Robert P. Russell (Washington: Catholic University of America Press, 1968), 146.
[9] Hostler, 31.
[10] Ibid.
[11] Ibid., 18-9.
[12] Ibid., 19.
[13] Ibid., 18.

errors," is a self-deception that provides the legitimacy needed for the will to regard an apparent good as Good.[14]

Innocence and Accountability

I find that many differences in scholars' interpretations of Leibniz's moral philosophy can be attributed to the lack of focus on Leibniz's concept of judgment. I posit that the philosophical tenability of the Leibnizian tripod is threatened by the role of accountability, or lack thereof, set forth by the innocence tradition. In contrast, the passion tradition claims that it is possible to commit error while also having clear knowledge about what is Good.

Of the innocence tradition, Maitra interprets Leibniz's philosophy in such a way that error can only be committed unknowingly, and thus not *akratically*.[15] Maitra's reading finds that judgment can be false while still being representational of perceptions, which are always true. A false judgment is an error in which the "mistake is to believe (i.e., to judge as true) that which is false," namely, to believe an incorrect interpretation or inference.[16] Maitra explains that the representation of perceptions might not be clear or distinct, but that the intellect will try to comprehend these confused and incomplete perceptions.[17] The error of judgment then occurs in the privation of completeness, where one's limited understanding modifies confused and incomplete perceptions to find meaning.[18] What one immediately perceives is a truth, although the intellect may mistake this truth for a falsity. The intellect may mistake an actual Good for an apparent good and vice versa, which would lead to committing an unknowing error.

Through her interpretation of Leibniz, Maitra claims that he succeeds in maintaining the Goodness of God because perceptions, all of which are received from God, are true even though these perceptions may be misunderstood to be something they are not. Her use of human "error," however, is confusing as it is unclear whether it refers to what is morally wrong or just factually incorrect. Maitra makes no effort to consider *akrasia* in her work and finds that human error is the result of limited judgment leading to faulty understanding. By disregarding the possibility of *akrasia*, her interpretation of Leibniz's ethics finds the first pillar, the goodness of human beings, to mean that human beings are solely Good. As such, it would seem that human beings are morally responsible for acting according to factual misunderstandings. This regard for human innocence must inevitably lead to the conclusion that divine punishment is unjust, which threatens the third pillar of the tripod, God's Goodness.

Murray, unlike Maitra, finds a place for *akrasia* within Leibniz's moral philosophy. Murray, of the passion tradition, claims that Leibniz's philosophy

[14] Leibniz, "Discourse," §23.
[15] Keya Maitra, "Leibniz's Account of Error," *International Journal of Philosophical Studies* 10, no. 1 (2002).
[16] Ibid., 63.
[17] Ibid., 65.
[18] Ibid.

is consistent with the discipline of psychology in viewing the intellect as a deliberating faculty that chooses between courses of action.[19] When deliberation arrives at a judgment, the will follows an action that judgment has reasoned to be Good.[20] According to Murray's research, the psychological tradition finds that those courses of action judged to be Good could only incline, but not necessitate choice.[21] Murray asserts that this lack of necessitation is a symptom of *akrasia*.[22] He claims that passions are "appetitions resulting from unconscious or confused perceptions or apprehensions," such that *akrasia* occurs when unconscious drives influence volition.[23] In opposition to Maitra, Murray not only understands confused perceptions as passions, but employs *akrasia* to argue that "passions affect choice by changing the premises that the intellect employs in deducing the last practical judgment."[24] In a sense, Murray's understanding of the Leibnizian system involves possibilities of adjusting the playing field of deliberation in the interest of the passions. Thus, Murray finds that the agent does not realize what perceptions it is subject to during the deliberative process. Although he clearly understands the concept of judgment to involve complicated and unconscious factors, it is still not evident that Murray believes human beings may choose to commit error despite knowing the act in question is an error.

Unfortunately, despite Murray's recognition of the need for accountability in Leibniz's moral philosophy, Leibniz employed a different understanding of *akrasia* than Murray. The concept *akrasia* traditionally refers to the incontinent person who lacks self-restraint and control.[25] Murray is employing the traditional, Aristotelian notion of *akrasia*, and not the Leibnizian one by which human beings know better, but freely choose the apparent good over the real Good. According to Leibniz:

> It is a daily occurrence for men to act against what they know; they conceal it from themselves by turning their thoughts aside, so as to follow their passions. Otherwise we would not find people eating and drinking what they know will make them ill or even kill them.[26]

It is evident from this quote that Leibniz not only concerned himself with the problem of *akrasia*, but might have also conceived of a psychological

[19] Although Murray does consider some disparities within the psychology tradition, on this point he claims there is a general agreement among the consulted psychology faculty.
[20] Michael J. Murray, "Spontaneity and Freedom in Leibniz," in *Leibniz: Nature and Freedom,* eds. Donald Rutherford and J. A. Cover (New York: Oxford University Press), 114.
[21] Ibid.
[22] Ibid., 115.
[23] Ibid.
[24] Ibid., 118.
[25] Henry George Liddell and Robert Scott, *A Greek-English Lexicon* (Oxford: Clarendon Press, 1968), 54.
[26] G. W. Leibniz, *New Essays on Human Understanding*, eds. and trans. Peter Remnant and Jonathan Bennett (Cambridge: Cambridge University Press, 1996), 1xxxv.

maneuver of self-concealment of some of one's own mental processes. Under Leibnizian *akrasia* the person committing error is acting according to reason, renders what is chosen freely chosen. Even though the person might "not do so for the right reason," and thus not achieve the highest level of freedom from passions, he or she did make a choice based on reason. Thus the choice was made freely, and the person is therefore accountable.[27] It is not the passions that overtake the intellect, but rather, *akrasia* is when judgment misuses reason to self-manipulate the conception of what is Good so that the passions appear Good. I posit that those who often commit *akratic* acts are artisans of making their desires seem reasonable. Those who often employ *akrasia* as a way of defending their actions to themselves misuse reason in such a manner that allows them to will otherwise than the Good. This strategy leads the will to some misrepresented construction of goodness instead of the actual Good. *Akrasia*, then, is the manipulative misuse of reason against better judgment that allows for a final product, a volition, which chooses an apparent good.

The significance of acknowledging *akrasia* in Leibniz's philosophy is that it renders just any divine judgment on human behavior. This accountability of *akrasia* reconciles the third and second pillar of the Leibnizian tripod, that is, that God is just and human beings make errors, which can be done unknowingly, but which can also be *akratic*. The issue with the innocence and passion traditions is that human beings seem set up for inevitable failure for which they are being punished. Because the innocence tradition finds that knowledge of the Good is so difficult to grasp, it seems that human beings have no option aside from unintentionally committing error and, furthermore, no option but punishment. The passion tradition also finds a certain helplessness in human beings as it holds that error committed knowingly is a result of the passions' overwhelming influence on the unconscious. I find that both the traditions' interpretations of the second pillar—that human beings commit error solely because of limited knowledge (innocence tradition), and that passions unconsciously cloud judgment (passion tradition)—conflict with the third pillar in the Leibnizian tripod, that God is Good and punishes justly.

Understanding Judgment

A person's volition cannot be solely directed toward the Good if God justly punishes that person. In Leibniz's system, human beings have inherited some of God's perfections, but are separate from God because of their imperfections. Leibniz describes this distinct human agency as follows:

> We could call that which includes everything we express our essence or idea; since this expresses our union with God himself, it has not limits and nothing surpasses it. But that which is limited in us could be called our nature or our power.[28]

[27] Hostler, 32.
[28] Leibniz, "Discourse," §16.

For Leibniz, the limited aspects of human nature are due to a certain division from God, such that not all components in human beings are divine inheritances. Accountability and agency must be part of human nature instead of a divine inheritance. Otherwise, Leibniz would have to admit that God authors error and punishes unjustly. Judgment, then, must be part of what is limited: the nature of humanity, its agency.

I claim that the tenability of the Leibnizian tripod is successful on the basis that humans knowingly misuse reason. As I explained in the first section, Leibniz's conception of volition is that it is composed of both the will, *conatus*, and judgment, *opinio*. I posit that volition is the product of the relation between the will, which is a divine inheritance and therefore represents infinity, and judgment, which represents finitude. As such, the will (*conatus*) is part of the essence or idea of humankind that includes divine inheritance and expresses the union with God, and judgment (*opinio*) belongs to the privative part that is necessary for human agency. It is because *akrasia* is located in *opinio* that I can maintain that the human essence, indeed a word used to translate *conatus*, is Good and always wills what is Good. Therefore, Leibniz's distinction between the essence and nature of human beings makes it possible to account for the goodness of human beings as well as their accountability for error.

Although the capability of knowingly committing error initially appears to conflict with the pillar of human goodness, I have attempted to establish an interpretation of Leibniz's moral philosophy that coincides with his claim that "he who punishes those who have done as well as it was in their power to do, is unjust."[29] The will, which is a divine inheritance and Good, does not knowingly will error. However, it does have the metaphysical capacity of willing error and can be deceived to do so.[30] As such, the innocence tradition is correct in acknowledging the dangers of ignorance. Additionally, Maitra's argument that the intellect modifies perceptions to find meaning—creating falsity to make sense of truth—may be a way in which the psychological factors of the intellect may influence volition, similar to the misuse of reason. Maitra's tradition, however, does not consider the possibility of error despite knowledge of the Good. The passion tradition, on the other hand, finds that passions may determine the choice of the agent by "traversing" judgment, causing the agent to be "deceived by appearances of good."[31] The passion tradition does recognize an aspect of humanity that is not directed at the Good. It fails to acknowledge that human beings may freely choose actions based on reasons not according to the Good, but rather, reasons that appease passions and which render human beings accountable. Neither tradition explores how *akratic* judgment, i.e., free and conscious choice according to reason—correct or otherwise—renders divine punishment just.

By understanding *akrasia* as the intentional and free misuse of reason, Leibniz can claim that reason is tenable and infallible as long as one uses reason correctly. Leibniz's advice to humankind to not

[29] G. W. Leibniz, "Abridgement of the Theodicy," in *The Philosophical Works of Leibniz*, ed. George Martin Duncan (New Haven: Tuttle, Morehouse & Taylor, 1890), 201.
[30] Leibniz, "Discourse," §30.
[31] Joseph, 183.

commit error can be summed up by this statement: "His empire is that of reason: he has only to prepare himself in good time to resist his passions . . . "[32] In other words, the way to avoid a manipulation of reason is to use reason only to guide the will towards the Good—not to make passions appear reasonable. Leibniz is also able to render faith tenable because God always reasons correctly and thus is Good. Because of Leibniz's allusions to the psychological, such as the conscious and subconscious, it might be possible to suppose that Leibnizian judgment may conceal or exaggerate certain components of knowledge to the self.[33] Perhaps further research and analysis of Leibniz's psychological elements could improve an understanding of judgment that supports this articulated tenability of the Leibnizian tripod of human goodness, human error that is sometimes unintentional and sometimes *akratic*, and God's Good punishment.

Conclusion

By employing Leibnizian *akrasia* and the interpretation of judgment put forth in this work, human beings still retain inherent goodness as stated in the first pillar of the tripod, for the will is nevertheless directed towards the Good. Judgment can function *akratically* because judgment arises as a result of privation and not divine creation, which completes the second pillar of the tripod. As such, God's punishment can be just, which addresses the third pillar of the tripod and renders it tenable. Beyond the tripod, Leibniz is making a claim about human nature and the nature of reason. Human beings should not discard reason because it can be misused, intentionally or unintentionally, but rather, human beings should reflect on how they use reason and whether they misuse reason immorally.[34] It is not reason that is flawed, but rather, it is the limited nature of human beings that is the flaw. Perhaps Leibniz hoped that by practicing reason correctly, human beings might improve morally and further contribute to Goodness. ❖

[32] Ibid., 184.
[33] Murray, 114.
[34] Leibniz, "Discourse," §30.

Philosophy Comes Out of Lives

An Interview with Marilyn Frye

Marilyn Frye is a noted philosopher and feminist theorist whose works include *The Politics of Reality: Essays in Feminist Theory* and *Willful Virgin: Essays in Feminism* as well as various other essays and articles. Frye recently retired from teaching philosophy at Michigan State University. On February 26, 2013, the *Stance* staff met with Marilyn Frye to talk about her work, her life, and the status of women in the field of philosophy.

***Stance* (Ashli Godfrey): So what made you decide to become a philosopher?**

Marilyn Frye: My first thought is that I was "born this way." In a way, that seems right, because it just came to me so easily. But there is a story here. When I was an undergraduate, I was at Stanford. Freshmen were not allowed to take the introductory philosophy course at Stanford at that time. That gave the course a certain status because you were considered too young and immature to take it as a freshman. The women I was hanging out with in my dorm knew something about what philosophy was, had some idea. I didn't—I'd never heard of it. They had this course high on their list of things you had to do in order to be a serious person, an intellectual, a sophisticated person, and so I just kind of followed them. So I got into philosophy in the first place by peer pressure. Kind of an unusual way to go, I think.

It was a huge lecture class, probably three hundred or more students. I loved the course. I got really fascinated with things like Berkeley's idea that things don't exist except when they're being perceived. And then I think a thing that maybe got me, really hooked me, was how hard philosophy is. I think it was the first time in my educational experience that I had encountered anything that I really worked on and didn't get an A or a 4.0. I got a B+. I was like, what? This doesn't happen. I don't do my best and get a B+. It was a new thing in my life, and I loved it. I thought: "Okay, this is worth doing."

Philosophy has always been, I think, the hardest thing I do. It's just incredibly hard work. I'm very often writing about things I don't understand at all, and I'm just working to understand something, working to figure something out. In that way for me philosophy is a lot like various kinds of art and what artists do. You take on exactly what is problematic for you, you're at the outer edge of what you can figure out,

and that's what you take on. You self-set the problems and the challenges. I don't write about stuff I already think I know. It's not interesting at all. I write about stuff I don't get. I think that's one reason it's endlessly interesting to me. I'll hear somebody give a paper on something I know almost nothing about in philosophy, and I will just dive in there. That's why I'm a philosopher. I hope the rest of you will have that much fun and work that hard.

***Stance* (AG): Marriage equality seems to have become the focal point for much lesbian and gay political activism in the last decade or so. What are your thoughts on this, especially considering your views regarding heterosexual marriage?**

MF: You have correctly anticipated that I am not a fan of the institution of marriage. It's got a terrible history in patriarchy, and I know people manage to live with it in ways that are okay, but it's designed to live in ways that are not okay, it seems to me. I don't get the mindset, the habit of thought, thinking the state should be an agency whose actions make a relationship legitimate or real. People doing commitment ceremonies—that makes sense to me. Usually that's a matter of people doing some rituals that accomplish things in their social milieu. Birthday parties do that, you know? That's nothing I have anything against.

"Sometimes the people who hate and fear you are right about something."

As for the involvement of the state in legitimating relationships and the setting up of domestic lives, I don't see going to bat to get the state into my life that way, and it's sort of curious to me that so many lesbians and gay men want to do that. I can sort of figure it out, but it's not the way I look at things.

But then, the other thing is that, and this goes along with gays in the military, the gay movement (and they don't usually call it lesbian and gay because lesbians just get swept up under the male name of "gay") in its large public presence right now has only this focus. Many gay men and lesbians are doing a lot else that's not in the large public picture of what gays want these days. We want (supposedly) marriage and we want to be in the military. Great. Two of the greatest patriarchal institutions of all time [and] they're clamoring to get in. Not my politics at all, not how I look at things.

Now, I do recognize that in the present status quo, there are a lot of very concrete benefits to be gotten by getting these official statuses as married or in the military, such as pay and the possibility of various kinds of civil rights that should be there for everybody, [but] which are in fact limited to people in these statuses. So, it's not like I'd be out there

lobbying in any eager way against gay marriage. I'm not against people getting the full range of civil rights and so on. It's just that that whole politics of "I want into all this wonderful stuff you've got" just doesn't seem right to me.

But I have one small hope for that whole situation. There are some people out there who are very convinced that if same-sex pairs are allowed to marry, it will, in fact, undermine and eventually destroy the institution of heterosexual marriage. I hope maybe they're right. Maybe "gay marriage" is more destructive than I ever would have thought. Sometimes the people who hate and fear you are right about something. It's good to think about it. In that case, maybe all this stuff I don't think is good politics will, in fact, end up right where I want to go.

***Stance* (AG): What do you consider to be the biggest roadblock for the acceptance of lesbianism?**

MF: That question really puzzles me. Acceptance by whom? I mean, I already accept it. It's like accepting that the sun rises. What is this "accept" business? So, I'm not sure what you mean.

I'm not going to be liked by everyone. In the big picture, in the culture at large, I really don't need more than tolerance. I don't need acceptance by most people. I don't need to be embraced. Just put up with me, and I'll put up with you. There are a lot of people around that I don't think are people I would want for neighbors because of their politics and their views, their lives. But tolerance, I feel, is required. That is, until they're actually dangerous to me.

***Stance* (AG): Feminist philosophers are often critical of gender construction and the "gender regime." This creates interesting intersections with current movements of gender-identifying persons, such as the trans community and the "femme movement." Is this a place of conflict, and if so, can you expand on how current feminisms can structure a response?**

MF: I'm really sympathetic with both the poles you've sketched out there. I think that, as a feminist, you can see there are all kinds of things going on in the world that support the understanding that being a woman is something entirely defined as being the "other" of man, and as not having any positive content. And you also are very aware that there's this hierarchy of women and men. When you wake up as a feminist, you perceive all this, and you want to think, "No, this whole mode of doing things, of dividing people into these two, one against the other, needs to just be abolished. There shouldn't be any such thing."

One of the early writers about this was a French feminist named Monique Wittig. She argued that lesbians aren't women in just

this vein—that lesbians don't exist as the other of man. They exist—we exist—as something ourselves. And then she thought that if feminism were successful there wouldn't be any more women, because that category would just dissolve. There are other people who have thought along similar lines, along an analogy of slave and master. Woman, man. Slave, master. You don't just want to get better treatment for slaves—you want to abolish the whole thing, right? Let's just get rid of the whole operation, and then there will be no women (no slaves), and that's good.

But then there's another sort of moment in the feminist experience of life where being a woman is not, in fact, only being something relative to man. It is something you've lived, and you've adopted and taken on as something to be defended against oppression. You feel like you don't have to buy that woman is just the other of man. In fact, it may even seem like this whole idea that woman is just the other of man is nothing but a sort of self-serving male fantasy. He'd like it to be like that, but in fact women are something. We are something on our own. Woman as a positive category. We construct what woman is, thank you, by being it and by relations to each other and by chosen relationships, not imposed ones, with men. In that case, it seems entirely wrong to be thinking, "Well, let's just abolish gender—there won't be any women anymore." That's gynocide. And that's not at all where you want to go. What you want to do is make it more and more possible for more and more women to give the content to what it is to be a woman in many varied ways and in many varied situations. So these seem to me both pictures of what being a woman is that are both very available to you as a feminist.

Frankly, I'm prepared to just live with both of them and tolerate the ambiguity. There are these two lenses, and neither of them should be privileged as the only way to look at it. So I go with both, recognizing that they don't make sense in terms of each other, but they each make a lot of sense of my life. So, there it is. I think the ability to just hang in there with ambiguity is important for getting on with life. And important in politics so you don't keep generating these "all or nothing" ways of looking at things.

***Stance* (Esther Wolfe): I was wrestling with the ambiguity of it, and I think by posing the question I kind of wanted to be like, "Marilyn Frye, fix this for me!" And I really love that your answer embraced the liminality of those spaces and that you can live in that space.**

MF: They both help make sense of what's happening, so don't pitch either of them out. I think you can see in what I've written these two ways of looking at things emerging at different times. The second is articulated most fully in the one paper, "The Necessity of Differences." The other is not so fully articulated anywhere, but it's very present in the back of my mind in a lot of moments in *The Politics of Reality*. So

don't abandon either of those overall pictures, even though they don't fit together with each other. Good, I'm very happy to have been helpful to somebody interested in that question.

***Stance* (AG): How would you describe a "wild woman?" Would you consider this phrase to mean the rejection of patriarchal norms?**

MF: It's a trope. It's a figure of speech. It suggests an image of women in patriarchy being tamed and domesticated, and that you could escape this, the scene of your domestication, if you want. I think that's a very useful image in a lot of ways. In particular, I think of the institution of marriage and related institutions. There's a very wide phenomenon that when somebody gets old, their younger female relatives, usually a daughter, sometimes a niece, is the primary helper, caretaker, or supporter. That's a kind of domestication that's deeply entrenched in someone who devotes a huge amount of her life without any reward. That kind of service to somebody in the family or anything like it is a kind of domestication.

With any trope like that, there's really nothing you can bank on forever because in another frame of reference you realize that all humans are domesticated animals. That's called socialization, and you don't survive as a human being if you don't do that. So everybody's domesticated. But that's in one frame. In another frame, looking at some of these institutions, it seems like a useful image to think of women as domesticated even though we could be wild. Aother thing about the image of women as tamed or domesticated, is that it seems to carry the suggestion of a "before" when women weren't yet tamed or domesticated. When we talk about domestication, we're talking about going from wild to domestic.

Even if in the history of our species there was in fact a time when the way humans lived was not male domination, to think too much in terms of trying to figure a future that's like the past may not be the way to go. I would not want to get bogged down by following this trope too far, but it's cool for a while. Certainly, for some individual women, it has been the image of themselves in their present situation as domesticated and tamed. It's just about right for them, and it encourages them to escape the zoo and go out and be wild. It can be very supportive for interpreting, in a useful way, certain kinds of conditions that women live in fairly commonly in our culture.

***Stance* (AG): Is it possible for men to deconstruct their own masculinity in a manner similar to "whitely" people? If so, what might this process involve, and what might be the result?**

MF: I'm going to fuss about the word "deconstruct." Deconstruction, in its most philosophical meaning as given to us by Derrida, is a particular kind of linguistic analysis. I'm talking here about a real reconstruction

of a person and a subjectivity. Deconstruction sounds to me like writing an essay.

As a person who is white and very prone to whiteliness, I can certainly say that it is possible to reduce my whiteliness quotient a whole lot. I can become aware of it in ways that cause me to feel displeasure and discomfort. I don't want that. I don't want to be like that. I recognize it and I begin to get sensitive to it. So yeah, you can change. You can change pretty deep habits.

How men should do it, and whether it is concretely or really going to be possible for this man, that man, or many men, I don't know. That's not my problem. That's their problem. Take it on or not. My experience trying to take it on regarding whiteliness is that we can make some headway.

You also have to pick your friends. There are a lot of people that get very dissatisfied with you when you stop being the correct way, the way they want you to be. You get called a traitor or a lover of the sort of person they hate. You don't get called that stuff if you've got the right friends.

"One of the things about various kinds of activism . . . is you go into the belly of the beast and start criticizing, rejecting, and resisting. You find out how much they hate you."

If you go out in various, more challenging environments, you catch a lot of grief from people for it. That teaches you a lot, by the way. One of the things about various kinds of activism, and you see this in the work of Andrea Dworkin, is you go into the belly of the beast and start criticizing, rejecting, and resisting. You find out how much they hate you. "Oh, I didn't get that." You do find out something about racism and racists. If you are a white person, and you get less whitely. I think men who get less masculine, tone that down, shift away from it, and so on will often find out how much men hate women. So it's up to you. Do it or not.

***Stance* (AG): What would your response to the category skeptic be in terms of your view that social categories are necessary in order to be intelligible to others?**

MF: The category skeptic says that social categories are imaginary or fictional or unreal. I'm saying: they're real. When we socially make things, they're as real as houses, cars, and the other stuff we socially make. So, there is a distinction between real and fictional. The category *goth* is real. I think possibly the category *schizophrenics* is not. I think

there might not be a real thing there, though there is an effort to make there be one. But the category *Starfleet crew members*, that's fictional. There are fictional social categories, and then there are real ones. *Men*, *women*, those are real ones. *Athletes*, that's a real one. So I just think there is an idea that something which is socially made by us into our social relations can be made to go away. Just like somebody's imaginary friend. It's like you get talked out of it, and then it's not there anymore.

I think social categories are real. In fact, you can get hurt by them. They don't go away by wishing. It's as hard to get rid of as an old car. When does it finally rust out into dust? How hard do you have to work to get rid of that social construct? Well, similarly to get rid of *women*, to get rid of *addicts*. To have these not be categories? No. That's no easier than ridding the univers of materially real things.

Starting with Simone de Beauvoir, but in some other phases of feminist thought, theory, and philosophy, people have made the distinction between sex and gender very much with the idea in mind that sex is an unchanging given and gender is something that can change. I see where people thought that and, in fact, I think it was useful as a stage, and we should return to it for certain purposes sometimes. Along with that, though, came the image that we can just undo gender. It's merely social. It's not material. I see how it can seem like that, but it doesn't just go away. It is as real as the desk, the chair, and the car. It is going to be just as resistant to annihilation.

***Stance* (AG): The field of philosophy has a persistent reputation for being especially hostile and unwelcoming toward women. As a woman philosopher, can you comment on your own experience?**

MF: Very soon after I decided I was a philosophy major, I was explicitly told by a philosophy professor, in so many words, that philosophy is not a women's field and that women don't do well in philosophy. I think he meant to be helpful, that I shouldn't get into something that was not going to pay off for me.

So from then on, I was in a defensive posture. I've got to show them. I'm going to do this even though he thinks I can't—they think I can't. I took him to be expressing a general view, to be manifesting a general climate. I didn't take him to be speaking solely as just one individual person with a certain opinion. I later talked about our experiences with another woman who had been in that department at that time (as an instructor). She said in that time and place, she felt that if you were a woman and a philosopher, you were just a living, breathing contradiction. That expressed it very nicely, I think—the way it was for me for many years. I got started in this before the woman's movement really got cranked up. I was relatively unaware of and unaffected by it until I was already

a beginning professor. I really had to be in a stance of fighting to do this thing, fighting to prove I *could* do this thing. It tempts you also to some bad things, like thinking as you do succeed in doing it, "How exceptional I am. I'm a woman, but I'm exceptional, and I can do this." That's a very bad place to go, and so it can cultivate some vices as well as strengths and makes the whole thing harder, much harder.

Now I think it's more that hostility to women has gone underground. In this country and in this time, it would be very unlikely that this professor would tell this new female major, "This isn't a woman's field; you shouldn't try to do this," in so many words. But it's there. So, I think it may be even harder to cope. It was right up front for me. I knew exactly what I was up against. And I think it may be more confusing now, more surprising or shocking when a moment goes by when it suddenly is overt, and then it goes secret right away. You think, "Did he really say that?" or "Did I really pick that up?" Then you doubt yourself because you really can't quite believe that's actually there.

I think it may be even more problematic in many places. My general take on the question "Well, have things changed or not?" is it's a completely different world, and nothing has changed. There are many more women in philosophy. We are much more present and recognized. The Eastern Division just had a wonderful woman philosopher as its president. Linda Alcoff is a feminist philosopher, a really good philosopher, and she gets a lot of recognition. That wouldn't have happened thirty years ago. There were one or two women philosophers around when I was coming up, but just a tiny few, so things really have changed, they really have. Now, at least you know if somebody says something or does something sexist to you, you can call it. You can say, "That's sexist." I used to say that, and they said, "What's that?" Now they argue with me; they know what I'm saying.

***Stance* (AG): What advice do you have for an undergraduate who wishes to pursue a career in philosophy and do you have any advice specifically for undergraduate women?**

MF: Well, I guess if it's you, and you love it, don't get talked out of it. Just go for it. Historically, both men and women philosophers, artists, musicians, and so on, have often had to make their living at some day job or another. Maybe you don't get to be a philosophy professor. I think in most of the things you might want to do, things that are intellectually interesting, creative, and exciting, you will have your moments when somebody thinks a woman can't do it, and they are afraid she can. It's so threatening that a woman can do what they can do, and they have to get really mean. You'll run into bad stuff. But, if it's you, do it. Actually, that's not that different than my advice to men, except the women are going to have to be stronger to do it, and that's not fair.

Otherwise, do it feminist. Bring your life into it. Turn out philosophy that comes out of women's lives. Philosophy comes out of lives. It does not come out of cognition alone. Some people are going to head off into "What is it?" and "How does it work?" Other people are going to head off into "What does it mean?" That's me. Other people are going to go off into, "But is this good, is this right?" The people who are interested in meaning tend to divide up into those for whom issues about translating thought into language are most compelling, while others will be more in the other direction, translating printed word into action. Different things within these areas are what engage people, and that comes out of their life. It's stuff in their life that makes that compelling.

"I think in most of the things you might want to do, things that are intellectually interesting, creative, and exciting, you will have your moments when somebody thinks a woman can't do it, and they are afraid she can."

Women's lives and men's lives are different in many ways. If you try to do philosophy like you were living a man's life, it will take the vitality out of it. A lot of us tried to do that for a long time. It's painful, and it's not nearly as productive and fruitful as working with whatever is really coming out of your own situation and your own life. It will be coded. It's not like they could read back and tell you what your personality is out of what your interest is. It's not going to just be the topic; it's also going to be the way of working it and what other stuff you bring in. Philosophers need to take everything they know into their philosophy, including all that you are, in order to make it really fun. ❖