2019-2020

The UBC Journal of Philosophical Enquiries

The UBC Journal of Philosophical Enquiries

Volume 1, Issue 1
October 2020

The publication is produced in association with the UBC Philosophy Students' Association, supported by the UBC Department of Philosophy and the Arts Undergraduate Society (AUS) at the University of British Columbia, Vancouver.

Editorial Note

The production of philosophical knowledge is an inherently social affair. Indeed, one can characterize it, in the words of a renowned historian of science, as "produced by people with bodies, situated in time, space, culture, and society, and struggling for credibility and authority."[1] This is equally true for both senior scholars and undergraduate students. The importance of vibrant and stimulating student culture, especially outside the classroom, to the development of critical thinking skills should not be underestimated. The revival of the *UBC Journal of Philosophical Enquiries* last year (previously known as *Hemlock*) was conceived by the members of the UBC Philosophy Students' Association as a way to acknowledge and enhance the level of undergraduate academic and community engagement at the UBC Department of Philosophy.

It would be premature to talk about whether we managed to accomplish these goals or not. Our work has certainly been offset by the outbreak of COVID-19, which delayed the publication and introduced new challenges in envisioning the future of the journal. Nevertheless, I am pleased to finally present to the public the first issue of the *UBC Journal of Philosophical Enquiries*.

I am thankful to everyone who submitted their work for publication in our journal. We have received a staggering number of submissions from undergraduate students from Canada, the United States, and Singapore, which made the publication possible. While there might not be any one topic uniting every paper selected for publication, all of them demonstrate an exceptional ability for interpretative and conceptual clarification, philosophical writing, and critical enquiry. **Isaac Fairbairn** seeks to understand the conceptual relationship between criticism and freedom in late Foucault. **Logan Ginther** argues for the possibility of passive responses to aggression in Kantian ethics, despite the centrality of self-respect in Kant's deontology. **Chuxuan Liu** provides a timely analysis of the nature of democratic elections, emphasizing the epistemic virtues of participatory democracy. **Anna Stonehocker** engages with Thomas Metzinger's works on spiritual practice, providing a genealogical discussion of the dichotomy between spirituality and religion. **Leo Stevenson** suggests a functional argument in favour of the congruence model of personal authenticity. **Owen Crocker** shows the compatibility of divine omniscience and human freedom. **Trueman Andrews-Gibson** explores the moral side of our epistemic practices, arguing for the obligation to think about our beliefs critically. **Angelo Chan Borges** proposes to reconsider Aristotle's ethics through the discussion of self-indulgence and incontinence. Finally, **Christina Ge** examines the con-

1. This characterization belongs to Steven Shapin. Although Shapin applies it specifically to science, it is equally relevant when it comes to philosophy even if occasionally it is not as obvious. See Steven Shapin, *Never Pure: Historical Studies of Science as if It Was Produced by People with Bodies, Situated in Time, Space, Culture, and Society, and Struggling for Credibility and Authority* (Baltimore: The Johns Hopkins University Press, 2010).

ceptual framework of just sex; uncovering the power dynamics behind the notion of consensual sex, she introduces an alternative framework based on the notions of consent, desire, and negotiation.

To end, I would like to thank everyone who participated in the production of the first edition of the journal. I would like to express my gratitude to the editorial board for taking their time to read, select, and edit the articles for this issue as well as for their patience in regards to uncertainties that inevitably surround any new publication. This edition would not be possible without the support from the members of the UBC Department of Philosophy, who kindly agreed to provide editorial training as well as professional feedback for the selected articles. Finally, I would like to thank the members of the UBC Philosophy Students' Association for their unrelenting support.

Sincerely,

Vlada Asadulaeva
Editor-in-Chief

Contents

"We Have to Be at the Frontiers": Criticism and the Subject in Foucault's "What is Enlightenment?"

Isaac Fairbairn
University of British Columbia

Abstract

In his final lecture "What is Enlightenment?", Michel Foucault unexpectedly defines his philosophical project with respect to the Kantian and Enlightenment traditions that constituted career-long targets of criticism. I argue that Foucault's late turn to the possibility of individual autonomy in a world of totalizing power relations depends upon the style or mode of criticism that his lecture both calls for and performs. I first introduce Foucault's style and defend his lecture against a general change of incoherence by examining how he performatively criticizes his own philosophical and historical position. I then consider how this style is seemingly guided by the Enlightenment values of autonomy and freedom it disavows and demonstrate it is more productively understood as a form of nonpositive affirmation with reference to Foucault's preface to Bataille's oeuvre. Finally, I argue that the problematically individualistic and privatized nature of the autonomy Foucault proposes is a result of his radical redefinition of freedom as a perpetually partial and transgressive attitude guided by only the vulnerability of its immediate limits.

Written a year before his death, Michel Foucault's 1983 lecture "What is Enlightenment?" signals an unexpected and remarkable change in his attitude. Often contrasted with his scathing critique of the analytic tradition—as in his 1966 work *The Order of Things: An Archeology of the Human Sciences*—and exposure of the disciplinary mechanisms and docile bodies underlying the Enlightenment's promise of progress and autonomy, Foucault's lecture sees him ambivalently align his philosophical project with Immanuel Kant and the Enlightenment. However idiosyncratic Foucault's interpretations of Kant and the Enlightenment are, the risk he takes in invoking them indicates the significance of this lecture for understanding the nature of criticism and subjectivity in the world of totalizing power and knowledge systems that Foucault's earlier work

was instrumental in discerning. I argue that an appreciation of the style or mode of criticism that Foucault both calls for and performs is necessary to productively engage with his critique of Kant and the Enlightenment as well as the nature of the freedom opened up. I first explain this style and defend Foucault's lecture against a general charge of incoherence by considering how Foucault performatively criticizes his own philosophical and historical position. I then examine how this style is guided by the Enlightenment normative values of autonomy and freedom it disavows and demonstrate – with reference to Foucault's preface to Bataille's oeuvre—that it requires an understanding of the criticism Foucault enacts and delineates as a form of nonpositive affirmation. That is, Foucault gestures toward an understanding of criticism and freedom as a perpetually partial transgressive attitude which is guided by only the vulnerability of its immediate limits.

The most pressing issue at stake in "What is Enlightenment?" is Foucault's decision to align himself with the Kantian and Enlightenment traditions he so vehemently opposed throughout his career. As Jürgen Habermas, one of Foucault's most vocal critics, formulates the problem: "How can Foucault's self-understanding as a thinker in the tradition of the Enlightenment be compatible with his unmistakable critique of precisely this form of knowledge, which is that of modernity?"[1] To first clarify the nature of Foucault's identification with Kant and the Enlightenment. In Kant's 1784 "Answering the Question: What is Enlightenment?" Foucault finds the first practice of a form of criticism. While previous philosophical reflections on the present understood it with respect to some ideal totality—as exhibiting the characteristics of a distinct world era or partaking in a historical teleology—Kant defines the Enlightenment "in an almost entirely negative way, as in *Ausgang*, an 'exit,' a 'way out'"[2] of our self-incurred immaturity, or the subjection of reason to authority. By conceiving of the Enlightenment as an indeterminate and ongoing process, without recourse to its positive and inherent features, Kant becomes the first philosopher to engage the present in its individuality, as a differential historical factor independent of past or future events. According to Foucault, modern philosophy is still struggling with the question raised by Kant's recognition of the contingency of the present: "What difference does today introduce with respect to yesterday?"[3]

After explaining the philosophical novelty of Kant's essay, Foucault claims it captures something about its historical moment that suggests modernity ought to be envisioned as an attitude, instead of a period of time considered linearly, preceded by an "archaic premodernity, and followed by an enigmatic and troubling 'postmodernity.'"[4] Modernity is not an epoch that develops or deviates

1. Jürgen Habermas, "Taking Aim at the Heart of the Present," in *Critique and Power: Recasting the Foucault/Habermas Debate*, ed. Kelly Michael (The MIT Press, 1994), 152.

2. Michel Foucault, "What is Enlightenment?," in *The Foucault Reader*, ed. Paul Rabinow (Pantheon Books, 1984), 2.

3. Foucault, "What is Enlightenment?," 2.

4. Foucault, "What is Enlightenment?," 6.

from the Enlightenment principles of the 18th century, but the individual's mode of relating to the present. The attitude of modernity is a critical impulse, the constant task to freely exercise reason, articulated by the "motto" or "instruction" underlying Kant's vision of the Enlightenment: "*Aude sapere*: 'dare to know,' 'have the courage, the audacity, to know.'"[5]

A cursory response to Habermas' charge of incoherence is that Foucault identifies with the critical mode of relating to the present that defines the modern philosopher and individual, rather than the particular aims and ideas of the Kantian and Enlightenment traditions in which it originated; "the thread that may connect us with the Enlightenment is not faithfulness to doctrinal elements, but rather the permanent reactivation of an attitude—that is, of a philosophical ethos that could be described as a permanent critique of our historical era."[6]

Foucault's Kantian and Enlightenment self-understanding may be qualified, limited to the critical ethos or "critical ontology of ourselves"[7] developed in the lecture's second half, but Habermas' charge of incoherence still stands. Why does Foucault risk undermining his philosophical legacy by elaborating this new form of criticism with respect to the Kantian and Enlightenment traditions decried throughout his body of work?

To answer this question requires first attending to the shift in Foucault's own philosophical trajectory taking place in his late writings. Foucault spent his career exposing the pretensions to universality and progress of contemporary discourses—such as those of sexuality, rationality and criminality—and their production and oppression of the modern subject. The critical ontology Foucault presents in his final lecture likewise seeks the contingent "in what is given to us as universal, necessary [and] obligatory."[8] But while Foucault's earlier work seems to undermine all talk of human agency in its depiction of the subject as a docile body punched out by power and knowledge regimes, "it is also clear, especially in his late writings [...] that he is not abandoning the idea that 'we constitute ourselves as subjects acting on others.'"[9] In his final lecture, Foucault maintains that the production of the subject by knowledge and power mechanisms does not foreclose the potential for freedom. And it is by critically attending to our historical moment that we are somehow afforded the possibility "of no longer being, doing, or thinking what we are, do, or think."[10]

The relationship between criticism and freedom for the late Foucault is apparent in the lecture's unexpected turn to French poet Baudelaire's aesthetics of existence, or dandyism, to illustrate the attitude of modernity. For Baude-

5. Foucault, "What is Enlightenment?," 3.
6. Foucault, "What is Enlightenment?," 8.
7. Foucault, "What is Enlightenment?," 12.
8. Foucault, "What is Enlightenment?," 11.
9. Richard Bernstein, *The New Constellation* (Cambridge, Mass: MIT Press, 1992), 154.
10. Foucault, "What is Enlightenment?," 11.

laire, modern man adopts an attitude toward the fleeting contingency of the present which "simultaneously respects the reality and violates it."[11] That is, modern man respects or recognizes the arbitrary "discontinuity"[12] of his historical moment—taking himself as the "object of a complex and difficult elaboration"[13]—and is thus faced with the perpetual "task of producing himself,"[14] with violating rather than liberating or discovering some essential human nature.

Foucault's reference to Baudelaire's dandyism signals his understanding of personal conduct as a possible site for freedom; the individual's mode of relating with the world and ability to critically question inherited historical influences. As Mark Bevir explains: Foucault's use of the term "aesthetic" escapes the conventional terms of moral debate with its focus on normative rules and pressures, expressing his "conviction that we can question identities and norms only from our particular location, not a neutral one, and only to create ourselves, not to discover ourselves."[15]

Foucault locates the potential for freedom in the individual's mode of relating to their constitutive historical moment, the ethos or style they choose to adopt. Returning to Habermas' charge of incoherence: the lecture's style, or *how* Foucault acknowledges his own situated philosophical and historical position, enacts or performs the liberating capacity of the critical attitude described.

The most striking feature of Foucault's style in his discussion of Kant's essay is the significance attributed to what he terms a "little text"[16] far from representative of the analytic tradition Kant's three *Critiques* are associated with founding. While Foucault's reading of Kant's essay is faithful, his claim that it lies at the root of the disparate elements of modern philosophy – "from Hegel through Nietzsche or Max Weber to Horkheimer or Habermas"[17] and presumably to his own work—is certainly not. Foucault's exaggeration of the relevance of Kant's essay is more than a playful revisionism and enacts the "experimental" component of Foucault's critical ontology, which opens up a realm of historical inquiry by putting "itself to the test of reality, of contemporary reality, both to grasp the points where change is possible and desirable, and to determine the precise form this change should take."[18]

Foucault's Kantian self-understanding realizes these two aims of experimental

11. Foucault, "What is Enlightenment?," 8.
12. Foucault, "What is Enlightenment?," 6.
13. Foucault, "What is Enlightenment?," 8.
14. Foucault, "What is Enlightenment?," 8.
15. Mark Bevir, "Foucault and Critique: Deploying Agency against Autonomy," *Political Theory* 27, no. 1 (1999): 77.
16. Foucault, "What is Enlightenment?," 6.
17. Foucault, "What is Enlightenment?," 1.
18. Foucault, "What is Enlightenment?," 12.

criticism, which correspond with the respect and violation of the present carried out in Baudelaire's attitude of modernity. First, Foucault concedes Kant's influence without contradicting his earlier works' appraisal of the blindness of Kant's philosophical project to the contingent character of truth-claim, which professes to transcend its historical conditions by positing "*a priori* categories which are deduced to constitute the consciousness of the human subject, as that which organizes perception as a timeless and universal structure."[19]

That is, Foucault argues that Kant's commitment to universal reason arose from his critical reflection on the present and negative definition of the Enlightenment as the moment humanity put its own reason to use, which required transcendental critique to "define the conditions under which the use of reason is legitimate in order to determine what can be known, what must be done, and what may be hoped"[20] and avoid falling prey to dogmatism or illusion. While the Kantian-analytic question was "that of knowing what limits knowledge had to renounce transgressing,"[21] critical philosophy replaces it with a positive one that "takes the form of a possible transgression"[22] by revealing the arbitrariness of these purportedly necessary limits. Each philosophical tradition originates in a recognition of the contingency of the present and a struggle to secure freedom through the exercise of reason. By experimentally identifying his criticism with Kant, Foucault reveals the possibility of change, the susceptibility of the limit separating analytic and critical philosophy to transgression.

Second, Foucault's experimental Kantian self-understanding violates or problematizes the philosophical influence conceded. Foucault is connected to Kant by a shared critical attitude of "analyzing and reflecting upon limits,"[23] the only difference being the negative or positive stance towards the limits exposed. Foucault applies this critical attitude to the Kantian tradition he received it from by positioning Kant's three *Critiques* and analytic project with respect to his reflection on the Enlightenment. If the three *Critiques* were "necessary"[24] at this particular historical moment to secure the boundaries of reason—"the critique is . . . the handbook of reason that has grown up in Enlightenment; and, conversely, the Enlightenment is the age of the critique"[25]—then the Kantian search for "formal structures with universal value"[26] is not itself necessary. Foucault violates the vulnerable limit of the Kantian project—exposing the historicized nature of its claim to universalism—by practising the very mode of

19. Mark Olssen, "Foucault and Critique: Kant, Humanism, and the Human Sciences," in *Future of Critical Theory: Dreams of Difference*, ed. Michael Peters (Rowman Littlefield Publishers, INC., 2003), 82.
20. Foucault, "What is Enlightenment?," 5.
21. Foucault, "What is Enlightenment?," 11.
22. Foucault, "What is Enlightenment?," 11.
23. Foucault, "What is Enlightenment?," 11.
24. Foucault, "What is Enlightenment?," 5.
25. Foucault, "What is Enlightenment?," 5.
26. Foucault, "What is Enlightenment?," 11.

critical reflection inherited from it. Foucault's reflection on Kant's philosophical legacy demonstrates that the possibility of change consists in the extent to which it determines modern philosophy, the self-historicizing critical attitude Foucault performs and describes. As such, Foucault's stylized Kantian self-understanding respects the reality of Kant's influence and takes the form of a possible transgression of this reality, implicitly issuing a "call for change among modern philosophers to consider that the dominant legacy from Kant need not be the only one."[27]

Foucault's Enlightenment self-understanding likewise respects and violates its influence. He accepts his historical situatedness by declining to take up a detached position: "We must try to proceed with the analysis of ourselves as beings who are historically determined, to a certain extent, by the Enlightenment."[28] While the extent of this determination is unclear—the Enlightenment is a set of complex historical processes including "elements of social transformation, types of political institution, forms of knowledge, projects of rationalization of knowledge and practices, technological mutations that are very difficult to sum up in a word"[29]—one of its results is the "certain manner of philosophizing"[30] practiced in Kant's essay. And, Foucault claims, this principle of critique is in fact "at the heart of the historical consciousness that the Enlightenment has of itself."[31] If the value of Kant's essay lies in its self-consciousness rather than historical accuracy—Foucault concludes his exegesis by claiming Kant's essay is not "capable of constituting an adequate description"[32] of the social, political and cultural transformations of the Enlightenment—then the heart of the historical consciousness the Enlightenment has of itself is self-consciousness.

In other words, the heart of the Enlightenment is not a particular idea or principle but self-awareness, the capacity to take one's self and historical moment as contingent and an object of criticism. The historical process of the Enlightenment is dependent upon the recognition of the contingency of self and the course of history, for striving to define and free ourselves is only possible when we can first recognize ourselves as self-conscious agents.

Seen in this light, Foucault's Enlightenment self-understanding is an example of what Benjamin Pryer term's Foucault's "active appropriation" by which he "shows that the reflection on the present that inaugurated modernity opens up to the very possibilities for criticism, and ultimately paves the way for abandonment, of the Enlightenment."[33] Foucault is not aligning himself with the

27. Christina Hendricks, "Foucault's Kantian Critique: Philosophy and the Present," *Philosophy and Social Criticism* 34, no. 4 (2008): 372.
28. Foucault, "What is Enlightenment?," 9.
29. Foucault, "What is Enlightenment?," 9.
30. Foucault, "What is Enlightenment?," 9.
31. Foucault, "What is Enlightenment?," 10.
32. Foucault, "What is Enlightenment?," 5.
33. Benjamin Pryor, "Foucault's Enlightened Reaction," *Human Studies* 25, no. 3 (2002):

Enlightenment, but showing how the self-historicization he is practicing is itself a product of the Enlightenment. And the individual's capacity to adopt a stance towards their historical position, to attain self-consciousness, constitutes their agential ability to question, violate, problematize and modify the normalizing forces that the historical transformations resulting from the Enlightenment self-consciousness have produced.

To fully resolve Habermas' charge of incoherence: the style of Foucault's Kantian and Enlightenment self-understanding performs the extent to which the individual's critical attitude is a space that is both historically determined and conducive to a freedom from this determination.

I will spend the rest of my paper addressing the issue arising from Foucault's stylistic self-fashioning notion of autonomy: its entanglement with the Enlightenment values purportedly moved beyond. To refute this criticism, I will read "What is Enlightenment?" with respect to Foucault's preface to Bataille's oeuvre, where this transgressive style is more clearly presented.

First, by drawing out the possibility of the abandonment of the Enlightenment, Foucault invokes the Enlightenment concept of autonomy. Christopher Norris argues that Foucault's late turn from perceiving truth-claims as a totally relativized "reflex product of the epistemic will-to-power"[34] to the espousal of a "private-estheticist creed"[35] generated a tension in his thought. Foucault recognized that "the truth-values of enlightened thought . . . cannot be abandoned" without also "renouncing any claim to promote or articulate the interests of justice, autonomy, and human emancipation."[36] "What is Enlightenment?" is host to this tension, resulting from Foucault's implicit recourse to Enlightenment ideals. Despite Foucault's insistence his ethos is not affirming an Enlightenment teleology or the "empty dream of freedom,"[37] it nonetheless results in the "permanent creation of ourselves in our autonomy"[38] and gives "new impetus . . . to the undefined work of freedom."[39]

Foucault in fact indicates his awareness of this tension in his lecture: the limit-attitude is prevented from falling into "disorder and contingency"[40] by its orientating "stakes": "How can the growth of capabilities be disconnected from the intensification of power relations?"[41] A critical ontology seeks to separate

318.

34. Christopher Norris, "'What is Enlightenment?': Kant according to Foucault," in *The Cambridge Companion to Foucault*, ed. Gary Gutting (Cambridge University Press, 2012), 178.
35. Norris, "'What is Enlightenment?': Kant according to Foucault," 184.
36. Norris, "'What is Enlightenment?': Kant according to Foucault," 184.
37. Foucault, "What is Enlightenment?," 12.
38. Foucault, "What is Enlightenment?," 10.
39. Foucault, "What is Enlightenment?," 11.
40. Foucault, "What is Enlightenment?," 12.
41. Foucault, "What is Enlightenment?," 13.

the struggle for freedom and acquisition of capabilities from the Enlightenment ideals or "procedures of normalization"[42] that have given rise to restrictive disciplinary mechanisms. To do so, Foucault proposes the study of the concrete activities of man: the "forms of rationality that organize their ways of doing things,"[43] their technological side; and the "freedom with which they act within these practical systems,"[44] their strategic side.

Consequently, Norris practices the very form of criticism that Foucault is renouncing by tasking him to produce the normative guiding values that a critical ontology functions independently of and in opposition to. I agree with David Hiley's claim that the issue between Foucault and his critics turns "not on the goal of enlightenment but on the nature of critique and the nature of autonomy achieved."[45] To engage with the nature of the criticism Foucault alludes to at the end of his lecture and how it disassociates itself from guiding normative values, it is helpful to refer Foucault's preface to Bataille's oeuvre "A Preface to Transgression." There, Foucault argues that transgression must be liberated from ethics and "the scandalous or subversive . . . [since] its role is to measure the excessive distance that it opens at the heart of the limit and to trace the flashing line that causes the limit to arise."[46] Rather than being a process for undermining or affirming values, transgression contains nothing negative nor positive—a "philosophy of nonaffirmation" or the mere "affirmation of division"[47]—and points nothing beyond itself other than the limit which it forces to "face the fact of its imminent disappearance . . . to experience its positive truth in its downward fall."[48] Individual freedom within the godless and relativized world depicted in the "Preface" or the "disorder and contingency"[49] mentioned in Foucault's lecture is only orientated by the vulnerability of its immediate limit—the technological component of practical systems—and measured by the excess ("positive fall")—the systematic component—that its transgression occasions.

Norris's criticism, then, fails to engage the constitutive partiality or nonaffirmative nature of Foucault's limit-attitude. In opposition to Norris's argument, Foucault further marks his departure from Kantian and Enlightenment ideals in the lecture by practicing the transgressive style discussed. First, Foucault locates the technological and strategic aspect of Kant's practice: the totalizing forms of historical reflection and Kant's resistance to them through the negative

42. Foucault, "What is Enlightenment?," 13.
43. Foucault, "What is Enlightenment?," 13.
44. Foucault, "What is Enlightenment?," 13.
45. David Hiley, "Foucault and the Question of Enlightenment," *Philosophy and Social Criticism* 11 (1985): 74.
46. Michel Foucault, "A Preface to Transgression," in *Language, Counter-Memory, Practice*, ed. Donald Bouchard (Cornell University Press, 1977), 35.
47. Foucault, "A Preface to Transgression," 36.
48. Foucault, "A Preface to Transgression," 34.
49. Foucault, "What is Enlightenment?," 12.

definition of the Enlightenment he develops. Kant's failure, then, is to answer "how the use of reason can take the public form that it requires, how the audacity to know can be exercised in broad daylight"[50] by proposing to Frederich II a "sort of contract,"[51] in which the implementation of the distinction between public/private uses of reason will be the best guarantee of obedience. Rather than explicitly denounce Kant's subjection of his critical ontology to a political condition (and limit), which would amount to the subversive or ethical form of criticism he opposes in the "Preface," Foucault instead exposes the limit of Kant's criticism by abruptly leaving for Baudelaire, who proposes a means for realizing this freedom without recourse to a normalizing condition: the individual's aesthetic self-fashioning. Again, Foucault suddenly moves on after exposing this limit—it can only be realized in a "different place . . . art"[52]—to elaborate this freedom in the more accessible Enlightenment ethos.

Crucially, Foucault's transgressive style, as a philosophy of nonaffirmation, is guided only by the knowledge of the vulnerability of the immediate limit, which informs his decision to not explicitly distinguish himself from Kant and the Enlightenment nor to produce a more rigorous analysis of what this agency would consist in (as Norris faults him for). Instead, Foucault's transgressive attitude is circular and perpetually partial: "We are always in the position of beginning again,"[53] constituting a "misshapen and craglike" philosophical language which "is folded back on a questioning of its limits."[54] Since a critical ontology is animated by the drive to expose the contingency of its immediate limit, it is always renewing itself because there is no ideal final ground it can lay claim to. Foucault's claim in his lecture that his form of criticism gives impetus to the "undefined work of freedom"[55] is therefore literal. Freedom is derived *secondarily* from the creative act itself: the reflective criticism is only the adoption of a mode of relating to the present, to which the promise of freedom remains undefined and incidental. This criticism affirms a limited or partial being which, unable to aim for the recovery of "an original soil which produced and which will naturally resolve every opposition,"[56] is formed only by the transgression of the limits that determine it.

To conclude, the note of uncertainty underlying Foucault's positive conception of agency is an important part of his message. A limit-attitude, and the nature of the agency it occasions, is always partial and takes the form of an open problem to be revisited and reconstituted since, as I have argued, it is grounded by only the vulnerability and potential transgression of the immediate limit. Foucault consequently ends his lecture by emphasizing the perpetual incompleteness of

50. Foucault, "What is Enlightenment?," 5.
51. Foucault, "What is Enlightenment?," 5.
52. Foucault, "What is Enlightenment?," 8.
53. Foucault, "What is Enlightenment?," 12.
54. Foucault, "A Preface to Transgression," 44.
55. Foucault, "What is Enlightenment?," 11.
56. Foucault, "A Preface to Transgression," 37.

the future of criticism: "This task requires a work on our limits, that is, a patient labour giving form to our impatience for liberty."[57] To practise this limit attitude requires relinquishing the demand for total coherence driving some of the critics of the late Foucault I have discussed. Rather than seeking a radical new system, Foucault suggests that criticism as well as the freedom achieved by the subject practicing it is partial but nonetheless powerful in its ceaseless exposure of the contingent limits of the necessary. To realize this freedom, then, is to conceive of our critical work as a "small night lamp that flashes with a strange light, signalling the void from which it arises and to which it addresses everything it illuminates and touches."[58]

57. Foucault, "What is Enlightenment?," 15.
58. Foucault, "A Preface to Transgression," 44.

Bibliography

Bernstein, Richard. *The New Constellation.* Cambridge, Mass: MIT Press, 1992.

Bevir, Mark. "Foucault and Critique: Deploying Agency against Autonomy." *Political Theory* 27, no. 1 (1999): 65-84.

Foucault, Michel. "A Preface to Transgression." In *Language, Counter-Memory, Practice*, edited by Donald Bouchard, 29-52. Cornell University Press, 1977.

——. "What is Enlightenment?" In *The Foucault Reader*, edited by Paul Rabinow, 1-15. Pantheon Books, 1984.

Habermas, Jürgen. "Taking Aim at the Heart of the Present." In *Critique and Power: Recasting the Foucault/Habermas Debate*, edited by Michael Kelly, 149-157. The MIT Press, 1994.

Hendricks, Christina. "Foucault's Kantian Critique: Philosophy and the Present." *Philosophy and Social Criticism*, 34, no. 4 (2008): 357-382.

Hiley, David. "Foucault and the Question of Enlightenment." *Philosophy and Social Criticism* 11 (1985): 63-83.

Norris, Christopher. "'What is Enlightenment?': Kant according to Foucault." In the *The Cambridge Companion to Foucault*, edited by Gary Gutting, 159-196. Cambridge University Press, 2012.

Olssen, Mark. "Foucault and Critique: Kant, Humanism, and the Human Sciences." In *Futures of Critical Theory: Dreams of Difference*, edited by Michael Peters, 73-103. Rowman Littlefield Publishers, 2003.

Pryer, Benjamin. "Foucault's Enlightened Reaction." *Human Studies* 23, no. 3 (2002): 317-321.

On the Possibility of Passive Responses to Aggression in Kantian Ethics

Logan Carey Ginther
University of Saskatchewan

Abstract

In "On the Possibility of Passive Responses to Aggression in Kantian Ethics," I respond to Barbara Herman's suggestion that, on account of the primacy of (self-) respect in Kant's ethics, violent resistance to aggression could be required of agents in certain cases. Jens Timmermann's interpretation of Kantian imperfect duties and Kant's own concept of moral endowments jointly point to the possibility of simultaneous alternative opportunities to fulfill duties of love having sufficient moral weight for particular agents to outweigh the respect-based obligation Herman posits in such cases. Explaining in Kantian terms the moral content of a morally compelling, love-based decision for pacifism, I propose that the implications of allowing such a figure as a conceptual Kantian pacifist a place in Kant's ethics will not be all that objectionable, intuitively or from the Kantian perspective. Indeed, acknowledging the legitimacy of such love-based passive responses to aggression draws attention to some subtler forms of respect than have hitherto been widely remembered in Kant's ethics.

There are, unfortunately but understandably, many ways in which Kant's ethics can come across unappealing. This is perhaps especially true at only a passing glance; some of the familiar restrictions put in place by the famous universalizability test of the Categorical Imperative seem intuitively too strict. To take the obvious example: Kant controversially believed that no lie could ever be morally justified.[1] Even in some of its lesser-known subtleties Kant's ethics can be unhelpfully vague. Kant posits duties of virtue to make "*one's own perfection*

1. For an especially clear presentation of this, see Immanuel Kant, "On a Supposed Right to Lie from Philanthropy," in *Practical Philosophy*, trans. and ed. Mary Gregor (Cambridge: Cambridge University Press, 2008): 605-615, in which Kant claims that even a lie to a murderer at your door inquiring after your dear friend would be fundamentally immoral because it would fail to respect the murderer's rational agency.

and the happiness of others" one's ends,[2] but never delineates any precise rules for when and how often opportunities to act on these ends ought to be taken (nor which of these ends ought to take precedence in the case that opportunities to act on them compete in the practical world).

The focus of this paper will be Barbara Herman's suggestion that there is, at least in some cases, a specific Kantian requirement for violent self-defence.[3] Herman *forbids* passive acceptance of aggression in the name of the dignity of humanity, viewing aggression as a physical manifestation of lack of respect for rational agency. Herman's requirement strikes me as the sort of too-strict result which would be unappealing enough to some to turn them off Kantian ethics. I will show, however, that Kant possesses the conceptual tools to provide a clear answer to this worry, even without condemning Herman's ideal self-defender.

This paper will explore the possibility of grounding an option of pacifism alongside Herman's justification of violence in Kant's ethics. I will outline Herman's position and locate her requirement within Kant's system of perfect and imperfect duties; importantly, it can only fall into the latter category. Herman's ideal self-defender is receptive to a respect-based demand, but Kant's moral endowment of love will suggest an alternative Kantian personality sensitive to different demands of imperfect duty. By explaining decisions regarding imperfect duties with reference to moral endowments, I will show how Kant's system of virtue can accommodate a greater variety of character types than Kantians who emphasize only respect would suggest, and this will be significant not only in the case of violent versus nonviolent responses to aggression but perhaps also as an answer to the vagueness objection to Kantian duties of virtue more generally.

Barbara Herman's Kantian Self-Defence

Contemporary Kantian ethicist Barbara Herman has explored how maxims of violence are handled by the tests of the Categorical Imperative and reached what can generally be considered intuitively satisfying conclusions. Herman notes that acts of self-interested violence commit the fundamentally immoral "error of 'devaluing' or 'discounting' rational agency" in its human form.[4] This is an echo of the Kantian notion that, on account of their capacity for rationality, "[e]very human being has a legitimate claim to respect from his fellow human beings and is in turn bound to respect every other."[5] The problem with maxims

2. Immanuel Kant, *The Metaphysics of Morals*, rev. ed., trans. Mary Gregor, ed. Lara Denis (Cambridge: Cambridge University Press, 2017): 6:385-388. Following standard practice, specific references to Kant's work will use the Akademie pagination wherever possible.

3. Barbara Herman, "Murder and Mayhem: Violence and Kantian Casuistry," *The Monist* 71, no. 3 (1989): 411-431.

4. Herman, "Murder and Mayhem," 425.

5. Kant, *The Metaphysics of Morals*, 6:462. This is, of course, grounded in the Formula of Humanity to be found at Immanuel Kant, *Groundwork of the Metaphysics of Morals*, rev. ed., trans. and ed. Mary Gregor and Jens Timmermann (Cambridge: Cambridge University Press, 2012): 4:428-29.

of self-interested violence is that they refuse to acknowledge this fact. Violent actions performed in self-defence against an aggressor can be justified *if* they do not feature such maxims.[6] If "I am . . . acting . . . to resist the use of my agency (self) by another," Herman writes, the problematic part of the maxim—that it directly contradicts something rational agents must will—disappears, and the action can be permitted.[7] Indeed, Herman goes so far as to say that "[t]his gives [agents] more than permission . . . [I]t imposes a *requirement* that aggression be resisted."[8]

Note that, with this last claim, we are possibly departing from the realm of generally intuitively satisfying conclusions. Of course, I do not assume that what Herman means to imply is that *violence* is always required in response to aggression. However, it does seem as if there could be cases in which no other form of resistance would do and that in those cases a refusal to act violently would constitute a failure of self-respect unacceptable to Kantians (or, at least, to Herman).

While I agree with Herman's conclusions about when the use of violence could be *theoretically* justified, I find it difficult to believe that such a scenario could ever play out in the real world as perfectly as it would need to in order to be granted that justification. Admittedly, the quality of agents' maxims is often opaque to them,[9] even in situations that are not violently charged; it seems safe to assume that it would be even more difficult for agents to examine their motives in a state of fear. Perhaps an exceptionally morally self-aware Kantian agent could manage the necessary assurance that her maxim was strictly one designed to assert her respectability, but I doubt that the majority of us would (or should) be so confident in ourselves. While I think this is a promising place for pushback, it is not where I plan to question Herman's conclusion (though perhaps it will be motivation for the reader to seriously consider the alternative I plan to propose).

A more important point of contention is that I do not believe that *all agents* intuitively want to justify violent actions in the first place, never mind be required to perform them, even in the exceptional cases Herman describes. Because of this, what I take to be an obvious and legitimate subsection of (could-be Kantian) agents may be facing a serious concern if violence is (even only in some cases) an ethical *requirement*.

Imperfect Duties and Moral Endowments

Because her claim is referring itself to such a fundamental feature of Kant's

6. From here on, "aggressor" will signify specifically Herman's morally errant type.
7. Herman, "Murder and Mayhem," 427.
8. Herman, "Murder and Mayhem," 427. Emphasis added.
9. Kant himself observes this fact and remarks on the importance of fostering moral self-knowledge. See, respectively, Kant, *The Metaphysics of Morals*, 6:441 and 6:446-47.

ethics—respect for humanity—it may seem at first glance as if Herman is invoking a perfect duty; Kant himself considers respect to be owed to all agents and thus admitting of no exception other than as vice.[10] However, it would be a mistake to say that any *positive* instruction relating to the duty of respect could prescribe *particular actions*; "this duty will be expressed *only indirectly* (through the prohibition of its opposite)."[11] Thus, Herman can only be referring to an imperfect duty in requiring resistance to aggression. Perhaps there is a perfect duty never to be an aggressor; these are not, however, equivalent.

Imperfect duties are Kant's many (and occasionally competing) positive prescriptions.[12] The negative prescriptions of perfect duty[13] can command perfectly by being ever-present in the background as limiting conditions of action. Kant does not, however, "specify precisely in what way one is to act and how much one is to do" in the case of duties of virtue, instead allowing "a playroom (*latitudo*) for free choice" in their fulfillment.[14] This has led many to interpret these imperfect duties as strangely weak, optional, or otherwise less than fully necessary dimensions of virtue to be picked and chosen as an agent pleases.

In an attempt to save Kant from accusations of unhelpful vagueness,[15] Jens Timmermann rightly notes that, "Kant does not say that imperfect duties may sometimes be set aside in the interest of inclination"—he says only that perfect duties definitely may not.[16] Timmermann clarifies the imperfection of imperfect duties as consisting only in the fact that they can be "restricted by the demands" of other duties.[17] Thus, every time there arises for an agent an opportunity to fulfill an imperfect duty, that agent is faced with "an 'obligating reason' . . . which turns into an actual obligation *if and only if* it is not defeated by other, weightier moral reasons."[18] Opportunities to fulfill perfect duties would and, presumably, alternative opportunities to fulfill imperfect duties could[19] be such

10. Kant, *The Metaphysics of Morals*, 6:464.
11. Kant, *The Metaphysics of Morals*, 6:465.
12. Recall the examples offered in the introduction to this paper: ethical commitments to one's own perfection and the happiness of others. These general imperfect duties especially communicate the competition that can occasionally arise in Kant's system of virtue; *practically*, such commitments can be at odds with each other.
13. Recall the examples offered above, the strict requirement never to lie or be an aggressor.
14. Kant, *The Metaphysics of Morals*, 6:390. It is a theoretical point being made in this passage, but it is intuitive, too. Again, recall the examples given in the introduction above, to make one's perfection and the happiness of others one's end; how the fulfillment of these imperfect duties will look for various particular agents will depend on contingent situational factors such as those agents' selves and the ends of the others around them. More on this later.
15. In, e.g., Jens Timmermann, "Autonomy, Progress and Virtue: Why Kant has Nothing to Fear from the Overdemandingness Objection," *Kantian Review* 23, no. 3 (2018): 379-397, and Jens Timmermann, "Good but Not Required?—Assessing the Demands of Kantian Ethics," *Journal of Moral Philosophy* 2, no. 1 (2005): 9-27.
16. Timmermann, "Good but Not Required?," 16.
17. Timmermann, "Good but Not Required?," 18.
18. Timmermann, "Autonomy, Progress and Virtue," 383. Emphasis added.
19. Admittedly, Timmermann himself does not appear to make this second, weaker claim.

weightier moral reasons.[20]

Thus far, it would seem as if our would-be Kantian pacifist is in a tough spot; there is perhaps less playroom in decisions regarding imperfect duties than might have been hoped for. Even though the duty Herman is expressing is only an imperfect one, it does not look like an agent could get out of acting violently toward an aggressor if that were the only possible form of resistance (at least, not simply for reasons like squeamishness). However, the language of weightier *moral* reasons provided by Timmermann does tell us what to look for in justifying passive responses to aggression as a legitimate option for a Kantian agent. A discussion of Kant's own concept of "moral endowments"[21] will be helpful in explaining how various obligations hold different weights for different agents, how instructions regarding "in what way one is to act and how much one is to do"[22] to fulfill duties of virtue are clarified for particular persons.

Kant identifies "*moral feeling, conscience, love* of one's neighbour, and *respect* for oneself (*self-esteem*)" as examples of such morally important character traits and explains them as "*subjective* conditions of receptiveness" to the commands of duty.[23] Importantly, such endowments are "natural predispositions";[24] thus, it is not a duty to have more or less of one or another of these. Rather, they are to be understood as making up the essential moral personalities of given agents. Agents possessing more or less of one or another of these endowments would *naturally* be led to more or less (or *different*) action fulfilling the corresponding duties of virtue, and granting this is how Kant makes the vagueness of imperfect duties seem not so vague from the perspectives of individual agents. An example will be helpful at this point.

Kant's own example is that an agent's decision not to take an opportunity to fulfill the duty of loving her neighbour generally (by loving, say, a stranger) could be justified *if* she did so because she simultaneously had an opportunity to fulfill the same duty of love specifically toward her parents (or, presumably, some other close relation) which she decided to take.[25] Timmermann explains that, though such options look like "equally good means or ways of doing one's duty" to

However, it strikes me as uncontroversial (perhaps uncontroversial enough to explain Timmermann's not bothering to make it explicitly).

20. Indeed, this interpretation seems to better square with Kant's own suggestion that it is grounds of obligation, not obligations themselves which have varying strengths and weights for agents (Kant, *The Metaphysics of Morals*, 6:224). Timmermann discusses this passage in Jens Timmermann, "Kantian Dilemmas? Moral Conflict in Kant's Ethical Theory," *Archiv für Geschichte der Philosophie* 95, no. 1 (2013): 36-64. Not surprisingly given his interpretation of imperfect duties, Timmermann reaches conclusions similar to those I will come to below; he does not, however, appeal to Kant's idea of moral endowments as I will, and I think that doing so is significantly helpful.

21. Kant, *The Metaphysics of Morals*, 6:399-403.

22. Kant, *The Metaphysics of Morals*, 6:390.

23. Kant, *The Metaphysics of Morals*, 6:399. Emphasis added.

24. Kant, *The Metaphysics of Morals*, 6:399.

25. Kant, *The Metaphysics of Morals*, 6:390.

someone outside of the situation, they often appear significantly different from each other to the agent considering them, on account of pertinent, personal "moral criteria."[26] Presumably, this specific agent's love of her neighbour as a part of her moral personality—in this case, one inclined toward family matters—acts as the moral weight that tips the scale and decides which obligation she must take the opportunity to fulfill. Importantly, it is easy to imagine a different agent in the same case opting to fulfill the opportunity to be beneficent to the stranger instead, knowing that her family will understand her excuse for being late (her moral reasoning, her assignment of moral weights).

Considering how well the idea of moral endowments handles the example of competing opportunities for beneficence, it seems likely that it will be the appropriate tool to use in addressing our questions about possible Kantian responses to aggression. In the case above, an agent's particular moral endowments dictated which of two opportunities to fulfill the *same* duty of virtue she ought to take. The case of violent versus nonviolent responses to aggression will be importantly different. The reason for this is that Herman's ideal case of justified violence is immune to accusations that in it the defender fails to respect the aggressor.[27] Thus, in order to justify an agent's choice of nonviolence, I suggest we look to the possibility of an opportunity to fulfill a *different* duty of virtue outweighing Herman's respect-based requirement of violent resistance to aggression—namely, a duty of *love*. Contrasting respect with love makes sense within the Kantian framework; the last two moral endowments Kant lists are *respect* for oneself and *love* of one's neighbour, and these correspond with the general division of duties of virtue into self-focussed and others-focussed duties.[28] The fact that there are separate moral endowments paired with each of these competing virtues suggests that there will be agents for whom others-focussed duties of love are more important than self-focussed duties of respect. In the next section I will flesh out the Kantian description of such a love-endowed agent and consider in what way I think she would be led to respond to aggression.

Grounding a Love-Based Pacifism in Kant's Ethics

The first duty of love Kant lists is the duty of beneficence.[29] Kantian beneficence is famously "not merely benevolence in *wishes*" but rather an "active, practical"

26. Timmerman, "Good but Not Required?", 21. Timmermann's examples of moral criteria are "debts of gratitude or prior arrangements"; such criteria are admittedly finer-grained than Kant's moral endowments, but sensitivity to these is doubtless determined by moral endowments. Moreover, they apply easily to Kant's example.

27. In Herman's ideal case, the defender's maxim asserts her own respectability without denying that of the aggressor; an attempt to show that there is a more aggressor-respecting alternative action would likely be fruitless.

28. This is the division cited in the introduction of this paper. Again, the two ends that are also duties are "*one's own perfection* and *the happiness of others*" (Kant, *The Metaphysics of Morals*, 6:385-388).

29. Kant, *The Metaphysics of Morals*, 6:452.

form of love.[30] Additionally, it is ideally *humble*.[31] The Kantian benefactor should view herself as "put under obligation" by the continuing needs of others and should "quietly . . . take on . . . the hardship" of "making the well-being and happiness of others [her] *end*."[32]

Consider a hypothetical extraordinarily beneficent agent who has, "without a great deal of fuss,"[33] devoted her life to unquestioningly[34] caring for others at her own expense and inconvenience. She was not well-known or rich when she began this moral work, and it has not brought her praise or fortune; on the contrary, her willingness to help has often caused her great emotional and financial difficulty. Nevertheless, she feels a personal requirement to act as she does. It would seem inappropriate not to call such an agent deeply virtuous—and, indeed, Kantians possess the conceptual vocabulary to describe her how we would hope in this case. Kantians could describe such an agent favourably by attributing her moral decisions to her incredible love for her neighbour, locating the reason for her particular brand of virtue within the framework of moral endowments by calling her primarily *love-endowed*. We can, as Kantians, understand and even commend this agent's continued humility and her readiness to suffer for the sake of others.

Let us extend and complicate the thought experiment by considering what would happen if this agent were attacked by an aggressor. If, as I suspect she might, this agent found it *personally impossible* to physically harm her aggressor, I am inclined to think that she would want the moral sentiment behind her decision to remain passive to be understood as a further expression of her self-sacrificial spirit. Such an action would, after all, be structured identically to her many previous beneficent actions; she would be gracefully accepting inconvenience for the sake of another.[35] However, I think that the general Kantian response to

30. Kant, *The Metaphysics of Morals, 6:452.*

31. Kant, *The Metaphysics of Morals*, 6:453. Kant praises particularly the rich agent who does not make a show of her beneficence and the not necessarily rich agent who silently accepts the burden of being beneficent.

32. Kant, *The Metaphysics of Morals*, 6:452-453. The first two quotes are from 6:453, the last from 6:452.

33. Timmermann, "Autonomy, Progress and Virtue," 391.

34. At some early stages of writing this essay I considered including an "almost" here in case the reader is moved by Kant's remarks about the possibility of moral problems associated with being *too* humble (Kant, *The Metaphysics of Morals*, 6:434-37). I decided against doing so because, as will become clear, I am not particularly moved by such remarks and I do not think that any Kantian need be.

35. It has been pointed out to me that this agent's response may be *too passive* to be considered a positive exercise of duty, especially of the duty of beneficence as Kant explains it—indeed, as I have already noted, Kant remarks that proper beneficence must be something active and practical, something *performed* (Kant, *The Metaphysics of Morals*, 6:452). The description of this agent's response as accepting inconvenience for the sake of another is important; it shows what I take to be the positive content of her action, the way in which she is not simply *doing nothing* (she is *being passive*, not simply *not being violent*). This positive content is admittedly very subtle, but it does strike me as something more active than Kantian benevolence in mere wishes.

such a decision (certainly Herman's) would be to explain it as a betrayal of a regrettable—perhaps even morally inappropriate—lack of self-respect.

It strikes me that there is no need for the Kantian evaluation of this agent to so diverge from the intuitive one at the point of her decision for a passive response to aggression. It would seem arbitrary to call such a characteristic (indeed, characteristically *virtuous*) action *vicious*, and yet the Kantian who would demand a certain necessary amount of self-respect on the part of all agents would perhaps be led to such a reaction—at least, such is the suggestion of Herman's *requirement*. The immediately following will consider in more detail the objection I anticipate Herman and similarly inclined Kantians voicing; I hope to show that the concessions which would have to be made in allowing such an agent her place in the Kantian ethical framework would not be all that objectionable, even to a Kantian.

Respect Reconsidered

The general Kantian emphasis on respect is perhaps appropriately placed (or, at least, *understandably* placed). It is, after all, the fact that "[h]umanity itself is a dignity"[36] worthy of respect that is the "conception of value at the heart of Kantian theory."[37] Thus, taking issue with an agent's allowing herself to be humiliated, hurt, or even killed in her opting for a passive response to aggression is perhaps appropriate (or, at least, understandable). It is tempting to simply reply that this objection ignores the fact that "*respect* for oneself (*self-esteem*)" is a moral endowment[38] and that these are not something for which an agent can be faulted for not having,[39] or, better yet, to cite the imperfection of positively expressed duties and the option of a love-based personal virtue built in to Kant's ethics. This may work[40]—however, there is another argument to be made which may be more compelling to the respect-favouring Kantian.

I do not take it to be obvious that there would be no respecting of the agent's self going on in the case described above. Such an agent would undeniably be incredibly morally self-aware, and it seems likely that she would have some sense of how morally impressive her actions are (and thus how *respectable* she is). If it is true that, on account of her particular moral endowments, such an agent perceives commands to live her life humbly, lovingly, and passively, then in her acceptance of such obligations—including this choice for a nonviolent response—she shows herself to be respectable by displaying her receptivity to

36. Kant, *The Metaphysics of Morals, 6:462.*
37. Herman, "Murder and Mayhem", 422.
38. Kant, *The Metaphysics of Morals*, 6:399.
39. At least, it would seem inappropriate to fault an agent for not having much of an inbuilt capacity for self-respect *in the same way* as she may be faulted for being actively disrespectful toward herself. Her lacking self-respect could perhaps be called morally unfortunate, but not much more.
40. Indeed, I think it does.

the moral law.[41] Indeed, it is easy to imagine such an agent possessing a quiet confidence regarding her decision, a "pleasure . . . from being aware that [her] actions are consistent with . . . the law of duty"[42] as it has been presented to her. In this quiet confidence and *transcendent* pleasure there would most certainly be a sort of assertion of the defender's self-respect, even if it is not visible to the aggressor.

I am aware that acceptance of this argument perhaps requires some concessions about the primacy of duties of respect in Kantian ethics (or, at least, an acknowledgement of some subtler ways of being respectful). However, I do not find such concessions to be intuitively objectionable or, in light of the explanatory work the concept of moral endowments can do, particularly un-Kantian. I hope to have shown here that "[t]here are duties which are"—*even if only*[43] *for specific agents*—"far greater than life, and which can often be fulfilled only by sacrificing life."[44]

Conclusion

I think it will be admitted that both agents considered in this essay—Herman's ideal self-defender and my own imagined passive philanthropist—are easy to sympathize with. It is not difficult to understand and appreciate the seriousness of their respective moral convictions, and it seems intuitive to say that both agents respond appropriately, each response being more or less appropriate only insofar as it agrees with the underlying moral personality of each agent.

It is a relief, then, that Kant's ethics need not posit a *requirement* of violent resistance to aggression in the same strict sense that Herman does. By prescribing both a self-respecting duty to perfect oneself and an others-loving duty to take up the ends of others and pairing a moral endowment with each side of this division, Kant allows the ethical outlooks of both Herman's ideal self-defender and my own imagined passive philanthropist within his system of virtue. There may be a Kantian requirement for violent resistance to aggression in certain cases for certain agents, but it is only an imperfect one selected for fulfillment on account of those agents' moral endowments—and if this is the case, there is an equally (*imperfectly*) strict Kantian requirement for passive resistance to aggression, too. At least to more passively inclined agents, this will be significant.

But, indeed, Kant likely vindicates *more* than these two ethical outlooks and

41. Respectability is an outward manifestation of an internal respect for the law (Kant, *The Metaphysics of Morals*, 6:464).

42. Kant, *The Metaphysics of Morals*, 6:399. This is a definition of moral feeling which is called the "subjective aspect" of respect for the law at 6:464 and is thus exceptionally appropriate here.

43. By including "only" I mean to emphasize that this need not be true for all agents or even all agents within the Kantian framework.

44. Immanuel Kant, *Lectures on Ethics*, trans. Louis Infield (Indianapolis: Hackett Publishing Company, 1980): 154.

posits more than these two opposing imperfect requirements; as noted in the introduction, Kantian virtue is fundamentally open-ended. However, the open-endedness here is better understood as making room for an appealing diversity of virtue than as being unhelpfully vague. By pairing moral endowments with duties of virtue, we remind ourselves that there is more to the Kantian ethical outlook than respect, and it is likely that this can be cashed out in many practical questions; our discussion of violence has been only one example of this. The interpretive move of pairing these two concepts, then, may be significant for the analysis of Kantian virtue more generally, beyond the specific question addressed in this essay. Thus, not only pacifists but also committed Kantians curious about how far Kant's ethics can be taken in the practical world should find this result compelling.[45]

45. I would like to acknowledge my debt of gratitude to those who reviewed this paper and offered many helpful suggestions and challenging criticisms (students and faculty members at the University of British Columbia and Mehdi Ebrahimpour, a colleague at the University of Saskatchewan). Additionally—and especially—I thank Dr. Emer O'Hagan at the University of Saskatchewan, immeasurably valuable teacher and discussion partner at nearly every stage of this paper's development.

Bibliography

Herman, Barbara. "Murder and Mayhem: Violence and Kantian Casuistry." *The Monist* 72, no. 3 (1989): 411-431.

Kant, Immanuel. *Groundwork of the Metaphysics of Morals.* Revised edition. Translated and edited by Mary Gregor and Jens Timmermann. Cambridge: Cambridge University Press, 2012.

——. *Lectures on Ethics.* Translated by Louis Infield. Indianapolis: Hackett Publishing Company, 1980.

——. *The Metaphysics of Morals.* Revised edition. Translated by Mary Gregor. Edited by Lara Denis. Cambridge: Cambridge University Press, 2017.

——. "On a Supposed Right to Lie from Philanthropy." In *Practical Philosophy*, translated and edited by Mary Gregor, 605-615. Cambridge: Cambridge University Press, 2008.

Timmermann, Jens. "Autonomy, Progress and Virtue: Why Kant Has Nothing to Fear from the Overdemandingness Objection." *Kantian Review* 23, no. 3 (2018): 379-397.

——."Good but Not Required?—Assessing the Demands of Kantian Ethics." *Journal of Moral Philosophy* 2, no. 1 (2005): 9-27.

——."Kantian Dilemmas? Moral Conflict in Kant's Ethical Theory." *Archiv für Geschichte der Philosophie* 95, no. 1 (2013): 36-64.

Democracy or Epistocracy: Why Participation Matters

Chuxuan Liu
Cornell University

Abstract

With the Brexit referendum and Donald Trump's electoral success, the year 2016 makes scholars like Jason Brennan question whether democracy and the equal right to vote are of intrinsic value. In *Against Democracy*, Brennan contends that the "one person one vote" system of participatory democracy is too often corrupted by ignorance and misinformation, and hence should give way to a rule of knowledge, namely epistocracy. This essay attempts to rescue democracy by drawing from Mary Park Follett's model of a robust participatory democracy that emphasizes creative citizenship and integration. Follett's participatory democracy not only avoids illegitimate domination but also provides a proper civic education that resolves Brennan's concern.

Introduction

In 2016, democracy had a bad year.[1] With the Brexit referendum and Trump's success, scholars like Jason Brennan question whether democracy and the equal right to vote have any intrinsic value at all.[2] Are citizens qualified to make political decisions through democratic elections? More fundamentally, is the old-fashioned "one person one vote" system of participatory democracy justified? In *Against Democracy*, Brennan argues that voters' ignorance and misinformation lead to suboptimal and sometimes quite bad political outcomes in democracies, hence democracy should give way to a rule of knowledge, namely epistocracy, which empirically can generate better outcomes.[3]

This essay attempts to rescue democracy by drawing from Mary Park Follett's model of a robust participatory democracy that emphasizes creative citizenship and integration. I contend that Follett's participatory democracy might

1. Jason Brennan. *Against Democracy* (Princeton: Princeton University Press, 2016), vii.
2. Brennan, *Against Democracy*, ix.
3. Brennan, *Against Democracy*, ix.

provide a proper civic education to the general public that could resolve the ignorance issues Brennan raises, and therefore has the potential to produce equally good if not better policy outcomes compared with epistocracy. I further claim that Brennan is mistaken that political institutions are of instrumental value only. Analyzing Follett's conceptions of "power-with," I find the political decision-making process to be more justified in participatory democracy than in epistocracy because participatory democracy avoids illegitimate domination over citizens. As participatory democracy surpasses epistocracy both in terms of justified processes and substantive policy outcomes, this paper concludes that participatory democracy promises a superior form of government compared with epistocracy.

Roadmap

The rest of the paper will progress as follows. First, I will briefly sketch out Brennan's and Follett's central arguments, namely Brennan's moral competence principle and Follett's creative citizenship through neighbourhood politics. Second, I will evaluate the advantages and disadvantages of epistocracy and participatory democracy in terms of their capability to produce policies that better promote the general welfare. Specifically, I will focus on how epistocracy and participatory democracy shape citizens' incentive structures and provide citizens with different resources to become more politically competent. Third, I will turn to discuss why Brennan's instrumental understanding of government is incomplete and forward the argument that politics could empower people, therefore, process matters. Lastly, I will end with a closer examination of the practical difficulties Follett's robust participatory democracy faces, in particular, how to balance individuality and social unity in reality.

Brennan's Moral Competence Principle

Imagine you are a judge who is overseeing a jury in court. You observe that instead of a careful examination of evidence and deliberative discussion, this jury reaches a verdict ignorantly and irrationally. As a judge, do you have a moral obligation to intervene and override this verdict? Most of us would answer affirmatively. The verdict seems to be illegitimate and should not be enforced because this jury has failed its obligation as a vehicle by which justice is to be delivered.[4] Given the verdict can greatly affect the defendant's and others' life prospects, it seems unjust to enforce a verdict reached by an incompetent jury.

Brennan generalizes the moral significance of competence and applies it to governance.[5] The decisions of the state are often greatly impactful and can be imposed involuntarily on citizens through violence and threats of violence. In other words, policy decisions are essentially the same as verdicts. It follows that

4. Brennan, *Against Democracy*, 153.
5. Brennan, *Against Democracy*, 156.

policy makers should also be competent, i.e., informed and rational. According to Brennan, citizens have at least as strong grounds as defendants to expect competence from government decision makers as a matter of right.[6]

Are policy-makers in a democracy competent or not? In Brennan's opinion, representatives are not the only policy-makers. Since ordinary people elect their representatives, their votes are also considered to be significant political decisions and Brennan's concern is exactly with people's right to vote. Empirical evidence appears to suggest that the majority of the population are not competent enough to make important political decisions such as choosing their own representatives. To begin with, most citizens cannot identify any congressional candidates in their district during election years.[7] Citizens also generally do not known which party controls Congress.[8] For as long as political scientists have been measuring, average voters have been misinformed or ignorant about basic political information which often causes them to support suboptimal policies and candidates.[9]

In Brennan's view, the problem of voter incompetence is inherent to democracy. While on the surface, democracy seems to empower each individual citizen with an equal share of political power, this share is so small that individuals do not feel that their participation actually makes a difference. Therefore, democracy provides little incentive for citizens to use their political power with discretion, not to mention to actively inform themselves.[10] Brennan suggests that participatory democracy is dominated by two kinds of citizens: the Hobbits and the Hooligans. The Hobbits are the ones who are politically apathetic and have very low levels of political knowledge, whereas the Hooligans are people who are politically motivated but cherry-pick facts to strengthen their preexisting biases.[11] Although Brennan does not deny that there could also exist fair-minded, knowledgeable citizens, he argues that their voices will be overwhelmed by the ignorant and irrational masses.

Brennan's argument against democracy can be summarized like this: people are entitled to competent governance, but participatory democracy is overwhelmed with incompetent average voters; in order to counter-balance the political influence of the ignorant mass, we must reward the knowledgeable with extra political influences, i.e., a rule of epistocracy. Brennan's moral competence theory goes beyond the claim that competence is a necessary condition for just governance, but it is also a sufficient condition. Government is viewed as only

6. Brennan, *Against Democracy*, 156.
7. Russell Hardin, *How Do You Know? The Economics of Ordinary Knowledge.* (Princeton: Princeton University Press, 2009), 60.
8. Ilya Somin, *Democracy and Political Ignorance: Why Smaller Government is Smarter* (Stanford: Stanford Law Books, 2013), 17-21.
9. Brennan, *Against Democracy*, ix.
10. Brennan, *Against Democracy*, 51-52.
11. Brennan, *Against Democracy*, 5.

a vehicle to deliver policies that best promote the general welfare. In Brennan's own words, whichever hammer that works the best is the one we should aspire to.[12]

Follett's Creative Citizenship and Integration

While Brennan searches for the best hammer, Follett thinks that democracy is more than a coordinating agency, but a great moral leader as well.[13] What is democracy in Follett's theory? According to Follett, democracy should not be simply the rule of the majority, instead, it should be the rule of "an interacting, interpermeating whole."[14] Creative citizenship and integrative thinking are the two central ideas in Follett's democracy. As she puts it, "Democracy is an infinitely including spirit . . . [It is] neither extending nor including merely, but creating wholes"[15] Follett's vision of democracy is a process that creates a social unity through citizens' active participation.

First and foremost, participatory democracy is about unity and wholeness. Follett does not envision democracy as a mechanical aggregation, i.e., a matter of counting votes.[16] She advocates for integrative thinking which aims not to compromise or concede when there is a difference, but to search for win-win solutions. Opposite to divergent thinking, which is a mindset based on scarcity and zero-sum relationships, integrative thinking fosters abundance through a mindset of possibility. It is a process by which many options are created to foster a new way of being in the world.[17] Consider an example of two people in a library room. One person wants the window open to get some fresh air and the other wants it closed to avoid the draft.[18] If one adopts divergent thinking, one of the two would have to concede and lose. However, Follett's integrative thinking suggests that instead of sacrificing someone's interest, they can open a window in the next room, bringing in fresh air without the draft.[19] To generalize Follett's integrative principle and apply it to politics, she calls upon us to overcome our fears and embrace our enemies—that is, all those who think and are different from us. The power of integrative thinking at the individual and collective levels advances a process whereby we engage and honor our differences

12. Brennan, *Against Democracy*, 14.
13. Mary Parker Follett, *The New State: Group Organization the Solution of Popular Government* (Mansfield Centre: Martino Publishing, 2016), 333.
14. Follett, *The New State*, 156.
15. Follett, *The New State*, 157.
16. Jane Mansbridge, "Mary Parker Follett: Feminist and Negotiator (Foreword)," in *The New State: Group Organization the Solution of Popular Government*, ed. Mary Follett (Penn: The Pennsylvania State University Press, 1998), xxiii.
17. Mary Follett, *Freedom and Coordination: Lectures in Business Organization*, ed. Lyndall Urwick (London: Management Publications Trust, Ltd., 1949), 34-46.
18. Roger Fisher, William Ury, and Bruce Patton, *Getting to Yes: Negotiating Agreement Without Giving In* (Harmondsworth: Penguin Books, 1986), 41.
19. Fisher, Ury, and Patton, *Getting to Yes*, 41.

not as conflicts but as opportunities for learning.[20] In short, Follett aspires to a collective will that respects and absorbs our differences and unites us as a whole.

Follett's integrative thinking might sound too good to be true, which leads to the second question: how can we overcome our differences and learn to think and act as a whole? Follett answers that we must "walk the talk"—that is, "living" a democratic life of ideas and experiences rather than a passive life of abstraction and entitlement.[21] Follett uses a religious parallel to explain this concept of creative citizenship: we are one with God not by prayer and communication alone, but by doing the Goddess every moment, so we are one with the state by actualizing the latent state at every instant of our lives.[22] Our relationship with the state should not be one of rights and duties, i.e., paying taxes and expecting good governance in return, as Brennan might envision. Citizenship should be an activity constantly exercised and democracy does not exist unless each person is taking their share in building the state-to-be.[23] State should not be a machine which functions on its own with a simple press of a button. On the contrary, state needs to be a common project that every citizen is constantly working on.

Where should we exercise citizenship? Follett argues that citizenship cannot be exercised or learned in abstract by taking a civic lesson. Citizenship must be exercised in neighborhoods where we can find a definite objective for individual responsibility.[24] According to Follett, neighborhood community could be the macro setting where the power is grown by individuals in relationships with each other and where democracy may be realized.[25] By participating in neighborhood organizations, individuals could relate to society better, seeing their civic duty as a specific, concrete thing taking definite shape for him.[26] Unlike Brennan's citizens who might rely on on the omnipotent government to bring justice to them, Follett's citizens would have a keen sense of personal responsibility, taking upon themselves the blame for their evils and the charge for their progress.[27] In short, Follett's participatory state could be achieved through the creative power of man as brought into visibility and actuality through his group life.[28] Neighborhood community should be seen as an ecological system, much like a garden, in which the various elements and citizens interact with each other, tend to each other, in relational processes to create something beautiful together.[29]

20. Gary S. Nelson, "Mary Parker Follett—Creativity and Democracy," *Human Service Organizations: Management, Leadership, & Governance* 41, no. 2 (2017): 178-185.
21. Nelson, "Mary Parker Follett," 182.
22. Follett, *The New State*, 335.
23. Follett, *The New State*, 335.
24. Follett, *The New State*, 340.
25. Nelson, "Mary Parker Follett," 183.
26. Follett, *The New State*, 340.
27. Follett, *The New State*, 341.
28. Follett, *The New State*, 342.
29. Nelson, "Mary Parker Follett," 183.

Follett's creative and integrative democracy is one where citizens embrace each other's differences and bring their unique contributions to the state through active participation in neighborhood communities. The process of democratic participation is considered as important, if not more important than the policy result. Follett's democracy emphasizes on both individuality and social unity. It aspires to a dynamic and inclusive policy that is created by unique individuals who strive for a common project.

Overcoming Ignorance: Neighborhood Education

As I have explained Brennan's and Follett's positions, I now turn to examine Brennan's most important challenge, namely that most citizens are too incompetent to govern. The obvious advantage of Brennan's position is that his ideal government would be comprised of the most knowledgeable elites who have the necessary expertise to govern. However, do elites have an absolute advantage over commoners in governance? I argue that Follett's creative state, if implemented, educates citizens through participation in the long run whereas Brennan's elite regime runs the risk of bias and misinformation.

What Is Political Knowledge?

Brennan maintains that political engagement in democracies is more likely to corrupt and stultify than to ennoble and educate people, turning a Hobbit into a Hooligan at best.[30] He cites Sarah Birch's research which reviews nearly every published paper on compulsory voting, a form of mandatory universal political participation, and concludes that it does not improve voters' knowledge.[31] Brennan argues that this is because when citizens start voting, it does not by itself cause them to take a greater interest in politics or learn more.[32] When participation is not compulsory, citizens who participate more actively demonstrate a higher level of political knowledge not because participation educates, but because of a selection effect—that is, more informed citizens tend to participate more. Brennan also denies that democratic deliberation can be educational. He cites political psychologist Tali Mendelberg who discovers an endless list of biases voters suffer from and concludes that when groups engage in discussion, they do not generate empathy or approach the discussion with a mind open to change.[33]

I do not contend that the most common forms of democratic participation nowa-

30. Brennan, *Against Democracy*, 55.

31. Sarah Birch, *Full Participation: A Comparative Study of Compulsory Voting* (Tokyo: United Nations University Press, 2009), 49-51; Birch, *Full Participation*, 57-67; Birch, *Full Participation*, 140.

32. Brennan, *Against Democracy*, 58.

33. Tali Mendelberg, "The Deliberate Citizen: Theory and Evidence," in *Research in Micropolitics: Political Decision Making, Deliberation and Participation*, ed. Michael X. Delli Carpini, Leonie Huddy, and Robert Shapiro (Greenwich: JAI Press, 2002), 180.

days are not educational as Brennan has illustrated empirically. Yet the specific kind of participation Follett forwards differs from compulsory voting or simple group discussion. The most knowledgeable political scientists are not necessarily the best policy-makers because they might not have the necessary local knowledge or an integrative mindset.

Unlike Brennan, who wants citizens to acquire more knowledge about politics, Follett emphasizes knowledge about participation. Follett describes neighborhood education as a process where citizens could be helped to discover their greatest ability, see the stimulus to apply, the path of approach, and that the constituents of their neighborhood should not merely serve but should serve in exactly that way which will best fit themselves into the community needs.[34] A polity needs not every citizen to be a political expert. It could function well with some of its citizens contributing their expertise in politics while others may contribute through other means. Follett draws an analogy to the system of war registration, where men and women record what they are best able to do, which might, through the medium of the neighbourhood group, be applied to the whole country.[35] A nurse may be defined as a Hobbit by Brennan for not knowing how many Senators there are. However, she might possess valuable information regarding the deficiency of the local healthcare system that a political scientist is not familiar with. In Brennan's epistocracy, this nurse would undoubtedly be disqualified for political participation due to her ignorance in politics—and yet in Follett's democracy, she would be given an opportunity to learn how to pass on the information she possesses to those who might be better equipped to provide policy solutions. Brennan might be right that the current voting system is flawed because it does not extract the appropriate information from each person. With that being said, this flaw does not provide sufficient grounds for exclusion. Instead, we should aspire to Follett's model which could accommodate people's different endowments and enable citizens to contribute in their own unique ways.

Benevolent Elites?

Brennan is concerned with psychological biases but somehow does not seem to worry about the knowledgeable who might be subjected to biases as well. As he acknowledges, most citizens who are politically knowledgeable nonetheless process political knowledge in deeply biased, partisan, motivated ways rather than in dispassionate, rational ways.[36] This is particularly dangerous given that empirically political literacy is positively correlated with a college education, higher income, being a Republican, and negatively correlated with being Black or a female.[37] In the most ideal version of epistocracy, Brennan would

34. Follett, *The New State*, 339-340.
35. Follett, *The New State*, 340.
36. Brennan, *Against Democracy*, 37.
37. Brennan, *Against Democracy*, 33.

suffer from the loss of valuable local knowledge. However, a more realistic challenge is that Brennan's epistocracy is likely to give additional advantages to some of the already privileged social groups who are likely to abuse these advantages given their psychological biases. Since political knowledge is not evenly distributed among all demographic groups, epistocracy runs the risk of producing unfair policies that serve in the interests of the advantaged rather than the disadvantaged.[38]

Brennan tries to avoid this challenge by arguing that even in democracies, certain groups of people tend to do better than others and it is false to presuppose that voters will each vote for their self-interest only.[39] However, in the face of Follett's integrative democracy, Brennan's defenses seem feeble. In addition to local knowledge that elites may not possess, which is already discussed above, Brennan also needs to explain why citizens may not vote according to their self-interest in an epistocracy. In fact, Follett's citizens have a better chance to have equal considerations for other citizens' interest supposing that they possess an integrative mindset. Neighborhood organizations not only have the potential to enable us to discover our unique ability to contribute, but also could give us a definite objective for individual responsibility.[40] Citizens would not be able to understand their duty or perform their duty unless it is a duty to something.[41] The intimacy in neighborhoods might eliminate the vagueness of civic duty and therefore citizens may genuinely believe that they are the state, hence offers the possibility of taking on the responsibility to care for fellow citizens.

To clarify, I do not argue that Follett's citizens are immune to psychological biases. Instead, my claim is that Follett's participatory democracy has the potential to dilute biases by including more members of society and educating them to think in wholeness. Although Brennan's political elites may be more knowledgeable, they have little incentive to correct their own biases and care for the disadvantaged groups which are excluded from participation.

Why Are People Ignorant?

Despite the fact that Brennan's elites may lack local knowledge and be biased, Brennan is not mistaken that political literacy is an important part of good political participation. Does Follett's participatory democracy offer a solution for the issue of ignorance? I argue that it does, because it aspires to change the incentive structure that Brennan identifies as problematic.

Brennan states that most voters are not stupid: they just do not care.[42] Becoming politically literate requires at least some effort and citizens just find these

38. Brennan, *Against Democracy*, 227.
39. Brennan, *Against Democracy*, 227.
40. Follett, *The New State*, 340.
41. Follett, *The New State*, 340.
42. Brennan, *Against Democracy*, 30.

efforts unworthy of their energy and time. According to Brennan, the root of the problem is that citizens do not think that their participation matters. A vote makes a difference only if there is a tie; otherwise, it usually does not matter for whom someone votes or whether they vote at all. Yet the probability a person will break a tie is vanishingly small, even smaller than winning Powerball a few times in a row.[43]

Again, Brennan's facts are not false. However, his view is limited in the sense that he considers voting to be the only channel of participation. Follett's neighborhood political training directly deals with the incentive issue. When democracy is degraded to a matter of counting votes, citizens are bound to think of themselves as nothing more than a grain of sand in the desert.[44] Almost certainly, if a person thinks their share in the collective responsibility as a 1/500,000 part of the whole responsibility, that is too small a part to interest them to perform such an infinitesimal duty.[45] The neighborhood organization is designed to change this thinking. As Follett puts it, a person is not a grain of sand in the desert, but rather a key of a piano in a neighborhood.[46] The value of the key is not in its being 1/56 of all the notes but in its infinite relations to all other notes. If that note is lacking, every other note loses its value.[47] Participation through neighborhood, if implemented, could instill a sense of unique responsibility in citizens which may stay with them as they participate in bigger political communities. When participation is not just about casting a vote that probably will not break the tie, but about playing a unique role in one's own community, citizens could be better incentivized to become politically literate.

In sum, Brennan's epistocracy promises less competent governance than Follett's participatory democracy for three reasons. First, Brennan neglects that non-experts may possess valuable local knowledge that cannot be collected if they are excluded from political participation whereas Follett aspires to teach citizens how to participate. Second, Brennan's epistocracy is likely to result in a rule of biased and privileged elites, who are unlikely to possess the same integrative mindset as Follett's citizens, which might lead to unfavorable policies for the already disadvantaged minority groups. Third, participating in Follett's neighborhood communities might make people aware of their individual importance to the collective. This change in incentive shows promise to solve the political illiteracy problem Brennan identifies in the long-run, whereas by excluding people from participation, Brennan essentially deprives them of both their incentive and resource to become more politically literate.

43. Brennan, *Against Democracy*, 31.
44. Follett, *The New State*, 340.
45. Follett, *The New State*, 340.
46. Follett, *The New State*, 340.
47. Follett, *The New State*, 340.

More Than a Hammer

In the previous sections, I have followed Brennan's framework, treating politics as nothing more than an instrument to produce policies that best promote the overall welfare. In this section, I turn to challenge this very assumption and argue that the form of democratic government has an intrinsic value, namely constructing and preserving a morally justified power structure, which Brennan's epistocracy does not recognize. Given that Brennan's epistocracy does not necessarily has any instrumental advantage, he needs to justify the inequality in political influences as well as elites' domination over others that epistocracy inevitably produces.

In a nutshell, Brennan's disregard for unjustified domination arises from his pessimism. He thinks that citizens' participation in politics can neither provide them with a sense of non-alienation, nor avoid domination. Before I present his objections, I will first clarify what is at stake here.

Thomas Christiano holds that political participation serves each person's fundamental interest in being at home in their social world, not alienated from it, and in being part of a social process in which their views are taken seriously.[48] Unequal political influences deprive citizens of this fundamental interest because they do not have equal resources to affect the outcomes of collective decision-making that determines laws imposed on all. In Brennan's epistocracy, citizens who are not politically literate enough will certainly feel alienated as the political decision-making process is monopolized by the elites.

In addition to the sense of alienation, inequality of political influences also raises the question of subjection. In Brennan's epistocracy, state authority will be centralized in the hands of a few political elites which means the mass is subjected to the elites. Daniel Viehoff proposes that such a political arrangement violates relational equality. He argues that it is an important feature of certain valuable relationships like citizenship that they require us to set aside—and not act on—unequal power advantages in shaping our interactions and the norms and expectations governing them.[49] To borrow an image from Rousseau: democracies have a special authority because they answer our need to replace "whatever physical inequality nature may have placed between men" with "a moral and legitimate equality," so that men, "while they may be unequal in force or in genius, all become equal by convention and by right."[50]

It is important to note that Brennan does not deny the value of non-alienation or non-subjection. His response is that democracy could not safeguard any

48. Thomas Christiano, "Debate: Estlund on Democratic Authority," *Journal of Political Philosophy* 17, no.2 (2009): 228-240.

49. Daniel Viehoff, "Democratic Equality and Political Authority," *Philosophy Public Affairs* 42, no. 4 (2014): 337-375.

50. Viehoff, "Democratic Equality," 352.

of these values. In response to Christiano's nonalienation argument, Brennan claims that democracies could never feel at home because our homes are niches. Most of us are at home in our homes because we may unilaterally shape our homes to reflect our preferences, something we cannot achieve in a broader political community because individual citizens are nearly powerless in them.[51] He compares participation in democracy to doing a wave at a sports game. The wave will happen with or without one's own participation, so citizens can only passively choose whether to accept and enjoy the outcome or not.[52] Brennan also claims that subjection happens in democracy as well. According to Brennan, empowering people like me to vote tends to protect me only if the people like me tend to vote in ways that protect me. My vote does not prevent others from dominating and exploiting me. What prevents me from being dominated is that other citizens either choose to restrain themselves or are in some way restrained.[53]

What underlies Brennan's pessimism is a misconception that individuals cannot make a difference in a larger community. Democracy is no better than epistocracy because it is just a tyranny of the majority versus a tyranny of the elite. My task then is to prove that Follett's conception of power-with sheds light on the significance of individuality in a large entity.

Follett defines Brennan's mistake as the fallacy of finals: a misbelief that the final moment in the process of decision-making is more important than any other moments.[54] The policy decision is not made when the president signs their name on the bill. Instead, the process begins when the first citizen gives their representative a phone call and expresses their concern with the issue. Follett asks us to recognize authority as a process in which every participant shares equal responsibility and power. She uses the term "power-with" to describe this power that is jointly developed, co-active, and non-coercive.[55] She quotes the president of a company who aims for "a kind of management with authority all down the line, as contrasted with management by edict from a central source. We are trying to teach our men what their jobs are, and then insist that they shall exercise the authority and responsibility which goes with that job instead of relying on the fellow above them."[56] When we think of this process in a political environment, we can easily see where Brennan goes wrong. He only focuses on the final vote that breaks a tie and the final signature that authorizes a law, but does not appreciate that each citizen, when participating in politics, is already exercising their share of authority and henceforth contributing to the final project. The Congressperson is not superior to an ordinary citizen just because their decision is higher above. People's positions in the decision-

51. Brennan, *Against Democracy*, 93.
52. Brennan, *Against Democracy*, 92.
53. Brennan, *Against Democracy*, 98.
54. Follett, *Freedom and Coordination*, 147.
55. Mansbridge, *Mary Parker Follett*, xviii.
56. Follett, *Freedom and Coordination*, 147.

making process are only reflective of their specialities and functions, instead of being indicators of superior power. Therefore, empowerment is not lost until an epistocracy effectively removes all the steps taken by ordinary citizens prior to politicians' final decision.

Non-alienation and non-subjection are likely to be protected in Follett's participatory democracy because individuals could potentially feel their impact on their neighborhood communities and exercise their own authority by participating. As local communities are much smaller in scale compared with the nation, individuals are more likely to appreciate their unique bonds with fellow community members and with the community as a whole, which would give them a feeling of being at home. In addition, Follett points out that collective decision-making process needs not to be coercive or dominating. When citizens actively partake in this process by performing their parts, they are jointly developing the power of the state instead of being subjected to an arbitrary power imposed over them. Democracy needs not to be a tyranny of majority. Instead, democracy could be a process where every individual's decision matters.

Discussion and Conclusion

This essay compares Jason Brennan's epistocracy and Mary Park Follett's participatory democracy and concludes that we should still aspire to Follett's participatory democracy because it is capable of delivering substantively good policy outcomes and constructing a more desirable power structure. While Brennan poses a valid empirical challenge that citizens are too ignorant to participate in politics, Follett's education through neighborhood communities, if implemented, could resolve this problem and lead to a more inclusive and fair government. Moreover, Follett sketches out a vision of jointly developed power which sheds light on how citizens may be free from subjection in a democracy.

Although, in my opinion, Follett's theory aspires to a more desirable form of government than Brennan's, she still leaves a few questions unanswered. Skeptical readers might wonder whether Follett's vision is only wishful thinking. To test whether participatory democracy can happen in real life, Maria Elias conducted a phenomenological research in a neighborhood of Akron, Ohio, a fairly diverse community where the neighbors themselves created various self-governing institutions in response to particular issues that strike the neighborhood such as economic development, safety, historical preservation, housing development, arts and recreation, planning and zoning, cleaning and trash pick-up, and the needs of children, the elderly, and others.[57] Throughout her interviews, Elias finds a robust sense of participation and integrative thinking in the community. One member reflects: "I realized that we have a great array of people in the neighborhood, and I realized how rich we are... you learn what other

57. Maria Vernoica Elias, "Governance from the Ground Up: Rediscovering Mary Marker Follett," *Public Administration and Management* 15, no. 1 (2010): 16-17.

people's needs are. As a group we need to do something about it. So being involved in the group has really opened my eyes to what the various issues are in the different constituencies. And it has made me very passionate about the neighborhood."[58] Another member pushes the idea of neighborhood association further: "The plan is to have a couple of streets take care of themselves, like Campinha (a neighborhood group). This is the model of what we would want to see in other parts of the neighborhoods... when you have groups like Campinha who make events for their street and the whole neighborhood... We need more streets that take ownership of their little corner of the neighborhood and work for its improvements. Miraflores (another neighborhood group) should be the umbrella for all the other little groups."[59] As Elias' research demonstrates, Follett's neighborhood governance model can thrive and expand in local communities.

With that being said, given that Follett aims to promote individuality and social unity at the same time, how exactly should we balance the two in practice, especially when the community is large? Mansbridge questions whether Follett's integrated solutions are always possible.[60] There is an undeniable benefit in keeping others' interests in mind when we make political decisions even when integration is not possible; however, Follett lacks a solution when there is no win-win solution. If we do encounter a zero-sum situation, should some of us concede voluntarily, letting others place their interests above us? If nobody wants to sacrifice, does the collective have to exert coercive power to make the final decision then? Brennan has an easy answer to these questions. He would suggest that the most competent politicians would make a decision on people's behalf. However, this option is certainly not available for Follett because of her commitment to equal consideration for all citizens.

In relation to this question of individual and collective conflict, Mansbirdge also raises concerns regarding Follett's language of "group," "harmony," "likemindedness," and "similarity" which seems to suggest the obliteration of differences.[61] It is ideal if diversity and differences are not to be homogenized in a melting pot, but used as a source of unity and a driver of innovation[62] However, there seems to exist a tension between social unity and individuality that might be magnified as the scale of polity increases. It is easier to discern the value of one's individuality in a small town of two thousand people, but how can we maintain our individuality and a belief in social unity in a nation of two million? Is preserving individuality possible in larger polities?

Follett's theory is valuable as it might restore our confidence in democracy in

58. Elias, "Governance", 21.
59. Elias, "Governance", 20.
60. Mansbridge, *Mary Parker Follett*, xxiv.
61. Mansbridge, *Mary Parker Follett*, xxv
62. Nelson, "Mary Parker Follett," 181.

this turbulent age. Her emphasis on local communities and integrative thinking provides an alternative route to an improved form of democracy in the future. Albeit there are still many questions to be answered regarding the practicality of her creative state, Follett makes a strong case that democracy should be the political arrangement to which we should aspire.

Bibliography

Birch, Sarah. *Full Participation: a Comparative Study of Compulsory Voting.* Tokyo: United Nations University Press, 2009.

Brennan, Jason. *Against Democracy.* Princeton, NJ: Princeton University Press, 2016.

Christiano, Thomas. "Debate: Estlund on Democratic Authority." *Journal of Political Philosophy* 17, no. 2 (2009): 228-240.

Elias, Maria Veronica. "Governance from the Ground Up: Rediscovering Mary Parker Follett." *Public Administration and Management* 15, no. 1 (2010): 9-45.

Fisher, Robert, William Ury, and Bruce Patton. *Getting to Yes: Negotiating Agreement without Giving In.* Harmondsworth: Penguin Books, 1986.

Follett, Mary Parker. *Freedom and Coordination: Lectures in Business Organization.* London: Management Publications Trust, 1949.

——-. *The New State: Group Organization the Solution of Popular Government* Mansfield Centre: Martino Publishing, 2016.

Hardin, Russell. *How Do You Know? The Economics of Ordinary Knowledge.* Princeton, NJ: Princeton University Press, 2009.

Mansbridge, Jane. "Mary Parker Follett: Feminist and Negotiator (Foreword)". In *The New State: Group Organization the Solution of Popular Government*, edited by Mary Follett. Penn: The Pennsylvania State University Press, 1998.

Mendelberg, Tali. "The Deliberate Citizen: Theory and Evidence." In *Research in Micropolitics: Political Decision Making, Deliberation and Participation*, edited by Michael X. Delli Carpini, Leonie Huddy, and Robert Shapiro, 151-193. Greenwich: JAI Press, 2002.

Nelson, Gary S. "Mary Parker Follett – Creativity and Democracy." *Human Service Organizations: Management, Leadership & Governance* 41, no. 2 (2017): 178-185.

Somin, Ilya. *Democracy and Political Ignorance: Why Smaller Government Is Smarter.* Stanford, CA: Stanford Law Books, 2013.

Viehoff, Daniel. "Democratic Equality and Political Authority." *Philosophy & Public Affairs* 42, no. 4 (2014): 337-375.

Making Sense of Spiritual Practice: A Critical Analysis of Thomas Metzinger's "Spirituality and Intellectual Honesty"

Anna Stonehocker
University of British Columbia

Abstract

This paper offers a critical evaluation of Thomas Metzinger's conception of spirituality without religion in his paper "Spirituality and Intellectual Honesty." Metzinger argues that spirituality is founded upon "intellectual honesty," the same normative principle that underlies science. Spiritual practice for Metzinger is aimed at ethical integrity through the pursuit of knowledge. He presents spirituality as the opposite of religion, which he argues is fundamentally dogmatic and fideistic, and therefore intellectually dishonest. While Metzinger provides a strong case for the link between spiritual practice and intellectual honesty, his claim that spirituality is the opposite of religion is problematic. Firstly, it disregards the role that religious tradition has historically played in providing a social framework of ideals to make spiritual practice meaningful. Secondly, his opposition relies on a narrow understanding of religion which focuses on belief and underestimates the importance of practice.

In "Spirituality and Intellectual Honesty," Thomas Metzinger develops a philosophical conception of secularized spirituality, according to which it is an "epistemic stance," a desire for knowledge pursued through inner practice.[1] The core of spirituality is intellectual honesty, an ethical ideal which he argues also underlies science. Furthermore, he claims that spirituality defined by intellectual honesty is properly understood as the opposite of religion, which is dogmatic and fideist, dogmatism being the thesis that "it is legitimate to hold on to a belief simply because one already has it" and fideism the view that "it is also completely legitimate to hold on to a belief when there are no good reasons or evidence in its favor, and even when faced with convincing counter arguments."[2] In contrast, spirituality is the pursuit of genuine self-knowledge or

1. Thomas Metzinger, "Spirituality and Intellectual Honesty," Self-published (2014): 25.
2. Metzinger, "Spirituality," 25.

"insight," a liberating knowledge "reflexively directed at the practitioner's own consciousness,"[3] developed through "some kind of mental training and practice, an inner form of virtue or self-refinement."[4] Metzinger's vision of spiritual practice founded upon intellectual honesty is compelling. However, I argue that he obscures the need for socially developed frameworks to make spiritual practice meaningful; that is, he does not show how an individual would recognize genuine insight without any recourse to external concepts or frameworks. Furthermore, because religion has historically provided these sociological frameworks that support spiritual practice, it is wrongheaded to construe spirituality as the opposite of religion. This opposition relies on a simplistic conception of religion which overemphasizes belief and overlooks the importance of practice.

I will begin by reconstructing Metzinger's argument before considering some of its shortcomings. Metzinger firstly defines spirituality and intellectual honesty respectively, then elaborates their conceptual connection. He traces the growing interest in spiritual practice separate from religious institutions in the West to after the Second World War.[5] The defining characteristic of this contemporary spirituality is being "concerned with practice and not with theory."[6] Where religion is based on dogmatic and fideistic belief, "the standpoint of pure faith,"[7] spirituality continually seeks to develop knowledge through practice. Metzinger thus posits that spirituality is the opposite of religion on epistemic grounds, for religion is essentially a systematic adherence to unjustified beliefs, the "deliberate cultivation of a delusional system,"[8] whereas spirituality is a practice motivated by "the unconditional desire for knowledge"[9] and therefore must be free of ideological commitments.[10] This ideologically-neutral spiritual practice for Metzinger consists in "direct experience" leading to insight about oneself and reality. The genuine self-knowledge sought by the spiritual practitioner is a part of what Metzinger considers ethical integrity. He elaborates this normative principle as "intellectual honesty."

Metzinger provides a straight-forward definition of intellectual honesty as "not being willing to lie to oneself."[11] He argues that it is a "normative idea" because it requires a moral commitment to truthfulness. He traces this to religion, specifically Locke's view that we have a religious duty to not only believe, but to know God's existence.[12] Metzinger takes the ethical valence of one's duty toward God to be unconditionally truthful in seeking knowledge and applies

3. Metzinger, "Spirituality," 6.
4. Metzinger, "Spirituality," 7.
5. Metzinger, "Spirituality," 5.
6. Metzinger, "Spirituality," 6.
7. Metzinger, "Spirituality," 14.
8. Metzinger, "Spirituality," 25.
9. Metzinger, "Spirituality," 27.
10. Metzinger, "Spirituality," 6.
11. Metzinger, "Spirituality," 10.
12. Metzinger, "Spirituality," 25.

it to the individual, as a duty to oneself. As such, while the moral status of intellectual honesty was based in from religion, in spirituality it becomes an internal commitment without the command of God; it constitutes the basis of personal integrity. This "secularized" pursuit of truth as an ethical principle underlies both spirituality and science. Evaluating Metzinger's argument, I will first analyze his conception of spiritual practice, uncovering how it implicitly relies on religious tradition. I will then point to limitations in his definition of religion underlying his opposition of spirituality and religion.

Metzinger states that spiritual practice is "radically individual"[13] and aims at "direct experience"[14] as a path to insight. It is individual because it requires "unconditional truthfulness and sincerity towards ourselves,"[15] a personal commitment for which membership in a group is not sufficient. Metzinger explains "direct experience" using the examples of Vipassana and Zen meditation, which aim at a particular form of self-knowledge, through the "systematic cultivation of particular altered states of consciousness."[16] It is noteworthy that he describes Vipassana as "largely ideologically neutral,"[17] a view that I will call into question. Metzinger discusses how, in Zen and Vipassana meditation, knowledge revealed through "direct experience" is not clearly defined by spiritual practitioners, but is generally non-propositional, and involves mental transformation, "attained by dissolving the subject-object structure and transcending the individual first-person perspective."[18] However, he anticipates the problem of evaluating the subjective experience of meditation. He asks: "What does real progress consist in, how could one detect it, and are there any criteria for distinguishing illusions, delusions, and self-deception from genuine?"[19] In the case of Vipassana, he cites the "classical answer" of how to judge genuine insight: "The criterion is ethical integrity, the sincere pursuit of a prosocial, ethically coherent way of life that is observable in a person's actions."[20] Metzinger thus suggests that the progress of a meditator's practice is judged by their ethical character. However, it is at this point that his characterization of spiritual experience as "radically individual" and "ideologically neutral" becomes problematic. By what standards is the individual judged for their ethical character? The only content Metzinger has provided at this stage for ethical integrity is intellectual honesty. Yet this does not help him here, for exactly how to recognize intellectual honesty or "insight" is the question at hand. Robert Sharf addresses this issue of evaluating spiritual experience in his discussion of the Vipassana tradition, concluding that "a meditative state or liberative experience is identified not on the basis of privileged personal access to its distinctive phenomenology,

13. Metzinger, "Spirituality," 25.
14. Metzinger, "Spirituality," 6.
15. Metzinger, "Spirituality," 25.
16. Metzinger, "Spirituality," 7.
17. Metzinger, "Spirituality," 6.
18. Metzinger, "Spirituality," 7.
19. Metzinger, "Spirituality," 7.
20. Metzinger, "Spirituality," 7.

but rather on the basis of *eminently public criteria*."[21] Such public criteria may include the lineage of the meditator and their specific training, as well as their behaviour.[22] Moreover, Sharf observes that "such judgments are inevitably predicated on prior ideological commitments."[23] Indeed, a Vipassana or Zen meditator makes sense of their meditative experience with the help of a community of spiritual practitioners and the framework of concepts provided in Buddhist scriptures and commentarial tradition, all of which involve ideological commitments. In this way, Metzinger's conception of direct experience in meditation ignores the role of ideological commitments within Buddhist traditions which have, over time, developed frameworks for recognizing and cultivating progress in spiritual practice.

Metzinger might respond that Buddhism indeed provides such frameworks, but should not be understood as a religion, precisely because it is based on individual spiritual practice rather than dogmatic belief. Indeed, many Buddhist modernists posit that Buddhism should not be seen as a religion because it does not require adherence to a set dogma. Evan Thompson addresses the trend in Buddhist modernism of divorcing Buddhist meditation practices from religious tradition. These meditation practices are nonetheless aimed toward a non-conceptual psychological shift, variably referred to as "awakening" or "Enlightenment." Describing the ongoing questions as to what "Enlightenment" means, including attempts to identify its neural correlates in the brain, Thompson argues that a commitment to Buddhist soteriology is necessary to make sense of the concept. He states that "if you accept the religious framework of Buddhism, which is fundamentally premised on faith in nirvana, you can give answers to these questions."[24] Whereas, "if you're outside the framework of that faith, it's not clear you can say anything meaningful."[25] Here Thompson highlights how the framework of faith is necessary for making sense of the mental transformation that meditation aims at. Thus, if we accept Thompson's claim, Metzinger's objection that Buddhism is not a religion overlooks the element of faith that guides Buddhist practice. Furthermore, because he does not provide examples of spiritual practice apart from those rooted in Buddhism, his argument for a general category of spiritual practice separate from religion lacks support.

At the heart of this debate is the question of how to define religion. This leads to my second critique of Metzinger's argument, that it relies on an overly simplified definition of religion. In his view, religion is fundamentally dogmatic

21. Robert H. Sharf, "The Rhetoric of Experience and the Study of Religion," *Journal of Consciousness* 7, no. 11-12 (2000): 279-280.
22. Sharf, "The Rhetoric of Experience," 279.
23. Sharf, "The Rhetoric of Experience," 279.
24. Evan Thompson, "The Rhetoric of Enlightenment," in *Why I Am Not a Buddhist* (Yale University Press, 2020), 157.
25. Thompson, "The Rhetoric of Enlightenment," 157.

and intellectually dishonest,[26] being "the standpoint of pure belief."[27] In *The Meaning of Belief*, Tim Crane argues that contemporary new atheists have misunderstood religion, reducing it to "a kind of primitive cosmology—a primitive or proto-scientific theory of the universe as a whole—or as simply a moral code, or as some combination of these two things."[28] Crane rejects this definition, and proposes a broader conception of religion as "a systematic and practical attempt by human beings to find meaning in the world and their place in it, in terms of their relationship to something transcendent."[29] He conceives of religion as most essentially defined by "religious impulse"[30] and practice. Religious impulse is the drive to connect to the transcendent, motivated by a sense that the world is imbued with meaning. As Crane states, "religion attempts to make sense of the world by seeing a kind of meaning or significance in things."[31] The second element, religious practice, is just as important as belief to religion, according to Crane, and involves a set of guidelines "about how to live one's life,"[32] not necessarily undertaken in relation to a supernatural agent.[33] According to a more nuanced definition of religion such as the one provided by Crane, faith-based belief may be a component, but is not the most defining characteristic or necessary to religion. Once the role of practice in religion is given more importance, Metzinger's conception of spirituality begins to seem much closer to religion, and certainly not its opposite. The stark opposition proposed by Metzinger between religion as intellectually dishonest and spirituality as intellectually honest does not stand once the definition of religion is broadened beyond faith-based belief. Furthermore, the dogmatic and fideistic thinking that his spirituality opposes is not only found within religion. Metzinger acknowledges this when he states that a significant part of contemporary "spiritual alternative culture" has become "intellectually dishonest and reactionary."[34] Metzinger should allow that spiritual practice aimed at intellectual honesty can occur within and outside of organized religion, just as dogmatic and fideistic thinking is found within religion and outside it.

My critique of Metzinger does not negate the possibility of a form of spiritual practice built upon intellectual honesty. Indeed, Metzinger provides a compelling case for the link between spiritual practice and intellectual honesty. However, my analysis of his conception of individual "direct experience" highlights the need for a sociological framework of ideals to make spiritual practice meaningful. Insofar as religious tradition has historically provided such

26. Metzinger, "Spirituality," 25.
27. Metzinger, "Spirituality," 25.
28. Tim Crane, *The Meaning of Belief* (Cambridge, Massachusetts: Harvard University Press, 2017), 3.
29. Crane, *Meaning*, 6.
30. Crane, *Meaning*, 50.
31. Crane, *Meaning*, 40.
32. Crane, *Meaning*, 82.
33. Crane, *Meaning*, 82.
34. Metzinger, "Spirituality," 23.

frameworks, it appears impossible to cleanly distinguish spiritual practice from religion. Metzinger's argument that they can be separated, and further, that spirituality is the opposite of religion, relies on a narrow definition of religion which disregards the importance of practice and overemphasizes the centrality of belief. Adopting a more nuanced definition of religion and deconstructing the opposition between spirituality and religion need not imply embracing dogmatic thinking or downplaying the importance of intellectual honesty.

Bibliography

Crane, Tim. *The Meaning of Belief.* Cambridge, Massachusetts: Harvard University Press, 2017.

Metzinger, Thomas. "Spirituality and Intellectual Honesty." Self-published, 2014.

Sharf, Robert H. "The Rhetoric of Experience and the Study of Religion." *Journal of Consciousness Studies* 7, no. 11-12 (2000): 267-287.

Thompson, Evan. "The Rhetoric of Enlightenment." In *Why I Am Not a Buddhist*, 140-164. Yale University Press, 2020.

Authenticity, Spontaneity, and Congruence

or

(What We Are Looking for When We Talk About) What We Talk About When We Talk About Authenticity

Leo Stevenson
Brown University

Abstract

In this essay, I outline two conceptions of what personal authenticity means. These are what I will refer to as the *spontaneity* model and the *congruence* model: the spontaneity model drawn from the work of Charles Larmore and Stephen Mitchell and the congruence model drawn from the work of Carl Rogers. In Section I, I articulate a basic sense of what a concept of personal authenticity does in light of a subjectively constituted understanding of the "self." On that basis, I establish grounds on which to compare competing models of authenticity. Section II details the two models and illustrates their consequences when applied, and Section III evaluates them on the basis of those consequences. My central claims claims are two: (a) given an understanding of the self as subjectively constituted, concepts of authenticity ought to be compared based on their consequences for what kind of self may be valorized as "authentic," and (b) because of its consequences, which facilitate better living, the congruence model of authenticity is preferable to the more traditional spontaneity model.

"*... to will to be that self which one truly is, is indeed the opposite of despair.*"[1]
— *Kierkegaard*, The Sickness Unto Death

In this essay I outline two conceptions of what personal authenticity means. These are what I will refer to as the *spontaneity* model and the *congruence*

1. Quoted in Carl Rogers, *On Becoming a Person: A Therapist's View of Psychotherapy* (Boston: Houghton Mifflin, 1961), 110.

model: the spontaneity model drawn from the work of Charles Larmore and Stephen Mitchell and the congruence model drawn from the work of Carl Rogers. In Section I, I articulate a basic sense of what a concept of personal authenticity does in light of an understanding of the "self" as subjectively constituted. On that basis, I establish grounds on which to compare competing models of authenticity. Section II details the two models and illustrates their consequences when applied, and Section III evaluates them on the basis of those consequences. My central claims claims are two: (a) given an understanding of the self as subjectively constituted, concepts of authenticity ought to be compared based on their consequences for what kind of self may be valorized as "authentic," and (b) because of its consequences, which facilitate better living, the congruence model of authenticity is preferable to the more traditional spontaneity model.

I. What We're Looking For

First, a few clarifying notes on my approach: on what we are doing when we call something "authentic," how an understanding of the self as subjectively constituted might change our understanding of authenticity, and the implications for how we ought to compare competing models.

Definitions of the term "authenticity" orbit around concepts of the genuine, the real, "truthful correspondence," sincerity, unaffectedness, correctness, veracity, verisimilitude.[2] I bring in the Oxford English Dictionary not to invoke any kind of sacred linguistic authority, but to back the general intuition that referring to something as authentic is saying, roughly, that it is "real" in some meaningful way.

The question then: "real" in what sense? Authenticity is not a question of something's "reality" per se—barring deeper metaphysical skepticism, an object is "real" in that it is there on the table in front of you, and an action or gesture (i.e., a behavior) is "real" in that it does in fact happen. Instead, the sense of being "real" in the concept of authenticity is a matter of whether something accords with the way it is represented. The descriptors "authentic" and "inauthentic," and any questions about their meaning, only become pertinent with regard to the conveyance of information. The authenticity of an object (e.g., a historical artifact, a work of art, or a drug) is tied to the object's actual origins or content in relation to their purported origins or content—is this papyrus really from Ptolemaic Egypt? Was this sculpture truly carved by the same man who carved Michelangelo's David? Does this pill really contain the chemical compound listed on its packaging?

The question of personal authenticity, then, is a question of whether one is "really" what one represents oneself to be. When we refer to behaviors[3] as au-

2. "Authenticity," OED Online. Accessed December 10, 2018. www.oed.com/view/Entry/13325
3. Also, activities, preferences, experiences, or mental states, though I will refer "behaviour"

thentic or inauthentic, we are referring to them as expressions of the "self" that convey information about a person to others. In cases where a behavior does not express information about the self, the descriptors authentic or inauthentic do not apply—for example, actions that are necessary rather than elective, like drinking water or using the bathroom, or that are especially mundane, like driving to work.[4] A claim about the authenticity of behaviors is entirely a claim of their veracity as expressions of the self.

A note: my claim here, namely that when we talk about personal authenticity we are referring to the veracity of behaviors as expressions of the self, only follows if we are treating the usage of "authenticity" that we employ when discussing personal authenticity (i.e., the authenticity of behaviors, etc.) as contiguous with the usage of the word that we employ when talking about other non-person things (such as the authenticity of works of art). This is a site for potential objections: some in the philosophy world view the word "authentic" when referring to persons as a fundamentally separate usage, possibly an esoteric one used primarily or exclusively in the world of academic philosophy. This brings up a key point of my approach. I contend, with the understanding that this position is debatable, that the point of a philosophical account is to clarify, elucidate, and possibly improve the foggy intuitional way that we use a concept such as authenticity. If that is the goal, then we may look to the common usage of the word to clarify its meaning: ergo I feel free here to define the "authenticity" in personal authenticity as contiguous with the concept of authenticity that refers to objects, meaning that it also refers to something's being in accord with how it is represented. To claim that the usage is fundamentally separate would be to abandon the task that our philosophical project set out to accomplish.

That task, to be clear, is the work of clarifying our concept of authenticity as it appears in our common usage, and comparing accounts of what it means to be personally authentic or to behave authentically. The goal of a philosophical account of authenticity is to judge what best delineates behaving authentically, "being oneself," from behaving inauthentically.

The question of what it means to be "oneself" or "true to one's self" is compli-

throughout this essay to avoid repeatedly listing the range of activities or qualities that can be authentic or inauthentic.

4. In Stephen A. Mitchell, *Hope and Death in Psychoanalysis* (New York: Basic Books, 1993), 130-131, Mitchell points to the example of "driving a car across a mountain pass on a rainy night" as an example of an activity where "self-expression has nothing to do with it," where asking about the action's authenticity is a meaningless question. Mitchell's example is incomplete, however, in that it omits the importance of choice in placing an action within the realm of (in)authenticity. Driving over that mountain pass would not be self-expressive if it were just what one had to do to get home; on the other hand, if the route were chosen for thrills or to impress someone, or the driving performed especially carelessly or timidly, the act would be self-expressive and therefore subject to the question of authenticity.

cated, however, by an understanding of the self as subjectively constituted. By "subjectively constituted," I mean simply that there is no objective, inner, true self as distinct from social influences or acts of interpretation; there is no "fact of the matter" or "true self" as a concrete referent to be represented truthfully. Subjective, interpretive processes of self-reflection or outside observation do not simply describe the self, but make it the way it is—they constitute it.[5]

Personal authenticity, then, is not an objective concept. An objective concept of authenticity would be like the authenticity of an object[6]—though it may be difficult or impossible to ascertain, there exists a definitive and straightforward answer to whether an object is "really" what it has been represented to be. A pill either does or does not contain a certain molecule in a certain quantity; a papyrus was grown, processed, and inscribed at some point in history; both either are or are not in accord with how they have been represented. There exists no straightforward criterion, however, for what distinguishes one's "inner" or "core" self from external influence. The self is fundamentally social: accreted over time, influenced by developmental surroundings, and variable based on its current context. The coherence of the self is a subjective construction. If personal authenticity were a question of expressing one's inner, original, objective "true self," then, the concept of authenticity would be meaningless—because there exists no such self to be expressed. Under such a model of authenticity, any outside influence would render a behavior inauthentic. Since such influence is always present, there would simply be no such thing as authenticity.

Rather than dismissing the concept of authenticity entirely, the thinkers I discuss in this essay replace the objective concept of authenticity with a subjective one. Subjective authenticity still describes accurate self-representation (even if only to oneself, as in the case of feeling that one is being authentic). But it differs from objective authenticity in that the self being represented is a subjectively constituted one—such that, in the case of "feeling authentic," the feeling of authenticity would be constitutive of the self's subjective coherence. The subjective concept of authenticity serves an interpretive function, then: with no "fact of the matter" of the self, calling behaviors more or less authentic is a process not of describing something inherent about them, but of determining their meaning to us. As a result, a different conception of authenticity will change the meaning that behaviors carry for us.

5. I understand that I am glossing this concept of the self rather too briefly. I make that choice partially because the concept is fleshed out a bit in the following paragraphs, and partially because there is too much subtlety and variation in literature on the self and social construction to unpack here. For my argument's sake, all that matters is that there is no clear, objective, realest thing that the self is—a thing which could be truthfully represented by authentic action. For more discussion of the self as a constructed entity, see e.g., Judith Butler, "Imitation and Gender Insubordination," in *Women, Knowledge, and Reality*, ed. Ann Garry and Marilyn Pearsall, (New York: Routledge, 1996), 371-387.

6. I bracket, for the time being, major epistemological questions about the possibility of objectivity.

Behaviors' meaning matters because conceptual distinctions between the authentic and the inauthentic always serve a particular function. For an artifact, for example, authenticity matters because it determines how much information the artifact might give us about its time period of origin; for a drug, authenticity determines whether its effects are likely to be what they claim. The possibility of inauthenticity becomes a concern where a false representation might lead us to a misunderstanding with practical impacts (e.g., the counterfeit medicine fails to properly treat the illness for which it is prescribed). Conceptual distinctions about personal authenticity will likewise serve a practical function; judgments of what is more or less authentic, especially given the value that our culture places on authenticity, are value-laden, to some extent prescriptive. Different conceptions of authenticity support different ideas of what expresses one's "authentic self" (even if that self is recognized as subjectively constituted), which in turn support different kinds of behavior. Our understanding of authenticity shapes the self, rather than simply describing it.

To summarize my groundwork so far: the common usage of the concept of authenticity indicates that authenticity refers to something's reality being in accord with the way that it has been represented to us. Since the self is subjectively constituted, personal authenticity cannot require an objectively accurate representation of a core "inner self." Personal authenticity instead functions as an interpretive, meaning-making concept, interpreting what representation of the subjective self is and is not accurate (and to an extent, though I did not dwell on this point, constituting the self that it claims to describe).

Moving on from that groundwork, we face the question of how to compare competing conceptual models of personal authenticity. Recognizing authenticity as an interpretive, meaning-making concept changes how we ought to compare those models. A more traditionally analytic approach might be to compare two conceptions in terms of their plausibility, examining which seems to better track the relevant phenomena, intuitions, and usage, with the goal of clarifying our concept of what we mean when we say that someone is being authentic. If personal authenticity is a matter of how we assign meaning to behaviors, however, it seems possible that two models may both be plausible. Rather than one being "more correct" than the other, they would just be different ways of interpreting the same phenomenon. Rather than seeking the most plausible concept, then, we might compare these models based on their function: we can aim to find a *preferable* model, based on the model's consequences.

I will hold for the purposes of this essay that the spontaneity model and the congruence model are both plausible, though making that case more thoroughly would take more space than I have. It seems to me that Rogers's congruence model is preferable, however, because in allowing the authentic self to be more self-aware and socially attuned, it supports preferable behavior.

II. What We Talk About When We Talk About Authenticity: Spontaneity, Congruence, and Their Applications

Spontaneity

The basic position of the spontaneity model is that to act authentically is to act unselfconsciously, without the conscious social interference that would make one "false" or "calculating." An individual acts inauthentically when she consciously adjusts her behavior based on social demands, consequences, or influences. One fails to act spontaneously, that is, when practical concerns force one to consider the way one's actions will be received by others. As Larmore (in his explication of Stendhal's concept of "le naturel") writes,

> Being ourselves is not . . . being true to an inner self that we already are; it is a quality simply of how we are at the time. Nor does authenticity require that we shake off the extent to which we have been influenced by others. Being essentially unreflective, authenticity entails only that we do not act with an eye to social convention—not that we are unaffected by it.[7]

That passage takes a view of authenticity as an in-the-moment quality of experience, as compatible with social influence, and as "essentially unreflective" (and therefore without consciousness of social convention). Mitchell, much like Larmore, writes that, "In differentiating authenticity from inauthenticity, the crucial difference lies not in the specific content of what I feel or do but in the relationship between what I feel and do and the spontaneous configuration and flow of my experience at that point in time."[8] Authentic feelings and actions do not have to be self-originating, that is, but have to be experienced in the moment as spontaneous. Authentic behaviors, according to Mitchell, "express spontaneous reactions to internal and external stimuli," while inauthentic behaviors are "contrived to manage self-states . . . and other people"[9] (e.g., acting in order to prove to oneself that one is a certain way, or to make a certain impression on someone). While Mitchell requires a lack of immediate self-consciousness and social contrivance, Larmore requires a more thoroughgoing lack of reflection. Reflection in general, writes Larmore, "operates with an eye to the anticipated reactions of others and so always shapes our behavior, to some extent, in accord with social expectations"[10]—whenever we are reflective, that is to say, we are necessarily acting based on social expectations. Larmore finishes the thought, writing that, "the distinctive thing about being natural

7. Charles Larmore, *The Romantic Legacy* (New York: Columbia University Press, 1996), 90.
8. Mitchell, *Hope and Death*, 130-131.
9. Mitchell, *Hope and Death*, 139-140.
10. Larmore, *The Romantic Legacy*, 89.

[i.e., authentic] is that it is unreflective."[11] Which is to say that any calculation of one's interests or motives, on Larmore's view, interferes with spontaneity enough to make an action inauthentic.

In short, the spontaneity model views authentic self-expression as necessarily composed of spontaneous action, which is to say action that lacks in-the-moment consideration of how actions will be received (or, possibly, lacks self-reflective consideration at any point). Recalling from section I that authenticity pertains to the accuracy of representation, we can understand this model as maintaining that the most truthful representation of the (subjectively constituted) self is that which emerges through an individual's unselfconscious, unreflective, spontaneous behavior.

Congruence

A congruence model, on the other hand, sees a representation of the self as most truthful when the individual is maximally aware of the content of her experience, feelings, and motivations and maximally able to express them to others. I draw the concept of congruence from the writings of psychologist Carl Rogers, who defines congruence as "a matching of experience, awareness, and communication."[12] Rogers notes that "when there is an incongruence between experience and awareness, it is usually spoken of as defensiveness, or denial to awareness. When the incongruence is between awareness and communication it is usually thought of as falseness or deceit."[13]

Rogers frames personal authenticity in the Kierkegaardian phrasing of "be[ing] that self which one truly is,"[14] an articulation that in itself is not different from that of the spontaneity model. The key difference is that Rogers believes that self-awareness facilitates authentic expression rather than impeding or preventing it. To illustrate incongruence, Rogers pictures a man who turns red and raises his voice in discussion, but claims "with evident sincerity and surprise, 'I'm not angry! I don't have any *feeling* about this at all!'"[15] This man seems to be acting spontaneously—he seems unselfconscious in the moment with regard to how his behavior will be received by others. His denial and surprise (provided they are truly sincere, as Rogers stipulates) show that he does not register his anger, as apparent as that anger is to those around him, and more importantly that he is not aware of *himself* as angry and denying it—if he were self-conscious, he would not shout so unconvincingly about how "not angry" he is. Despite his evident spontaneity, however, Rogers would say that in incongruently denying his clear anger he is not being fully authentic. "The self that

11. Larmore, *The Romantic Legacy*, 89
12. Carl Rogers, *On Becoming a Person: A Therapist's View of Psychotherapy* (Boston: Houghton Mifflin, 1961), 339.
13. Rogers, *On Becoming a Person*, 341.
14. Rogers, *On Becoming a Person*, 166.
15. Rogers, *On Becoming a Person*, 340.

he truly is" is different from the self that he is representing.

Rogers, like Larmore and Mitchell, approaches authenticity largely in the negative, as a meaningful lack of social interference: Rogers discusses clients who, in becoming the selves they truly are, move away from facades, from social demands or values that tell them what they ought to feel or be like, and from the need to meet expectations or please others.[16] Where Larmore and Mitchell place that social interference in the conscious moment of action, however, Rogers adds that it may also operate through a blanket defensiveness that operates on an unconscious level. Recall that the spontaneity model locates inauthenticity in the experience of self-consciousness: when the individual behaves in conscious accord with others' expectations or becomes aware of how an action will be received, that individual is no longer acting spontaneously, and therefore no longer acting authentically. The congruence model, in contrast, locates inauthenticity in the individual's defensiveness, which can create falseness either consciously or unconsciously. Rogers writes: "I quite literally cannot see, with accuracy, those experiences, feelings, reactions in myself which are significantly at variance with the picture of myself which I already possess."[17] When an individual adheres to a social identity tightly enough that they are unable to recognize the processes that truly make up their motivations and actions, their expression will be at odds with "the self which they truly are." As such, according to the congruence model, they will be acting inauthentically.

Applications

To further illustrate the difference between spontaneity and congruence, a couple of (fictional) illustrations of where the models would differ in their judgment of authenticity:

> (1) Home for the holidays, I find myself irritable and easily upset, and I berate my father for stomping up and down the hall while I am trying to read. If I took some time to reflect (or reflected afterwards, possibly in a therapeutic context), it would be apparent to me that my negative feelings are truly due to being in my parents' home: I grate at the lack of independence, or I notice old patterns emerging in myself and worry that I have not made as much progress in my personal growth as I had believed. Under a spontaneity view, I am being authentic when I berate my father: I do it without reflection or consideration of how this behavior will hurt him and strain our relationship. I am at that moment angry and want quiet, so expressing that unreflectively is authentically representing myself. Under a congruence view, I am being inauthentic: the self that I really am is not one who is angry at loud footsteps, but one who is upset by late-adolescent angst. In misidentifying the causes and dynamics of my negative emotion as I externalize it, I misrepresent

16. Rogers, *On Becoming a Person*, 167-170.
17. Rogers, *On Becoming a Person*, 187.

myself.

(2) Home for the holidays, I get in a fight with my father—he has been making unreasonable demands on me, badgering me about applying to law school, etc. and I decide that I cannot take it anymore. I have been discussing my feelings about my father in a therapeutic context all fall, and at this point I feel a full understanding of my anger towards him. I also understand that I am financially dependent on him, however, so I do not want to risk fully alienating him. When tempers reach a pitch I do express my rage, even remarking on my opinions of his character flaws. But I understand that I will be more likely to win our fight if I frame my anger using "I" statements, so in our shouting match I make declarations like "I feel terrible when you say that," and ask questions like "What do you want from me?" By not expressing my feelings through hurled insults, I strategically hold on to the moral high ground. Under a spontaneity view, I am being inauthentic: I am certainly reflective and conscious of the social impact of my presentation. My statements and behavior are calculated in the moment based on the impact they will have on my father, with the intention of emerging as the victor in our argument and influencing our future relationship. Under a congruence view, I am being authentic: I am aware of my feelings and motivations, and I am expressing them accurately. In fact, an unbridled and spontaneous expression of rage would likely be less authentic than a carefully worded attack. A statement like "you're such an idiot," for example, even if spontaneous as a speech act intended purely to wound, would not be an accurate expression of my experience.[18]

The significant difference that emerges here is each model's implicit view of the self that is being authentically or inauthentically expressed. Under a spontaneity model the self can be socially influenced, but its conscious experience in the moment has to be *as if* it were uninfluenced. The fallacy of objective authenticity as causal isolation is gone, but the subjective quality of authenticity is that of an individual who is momentarily consciously isolated. The congruence model, on the other hand, allows for an authentic self that is experienced as socially connected; thinking through one's impact on others is a part of one's experience, no less than the experience of an unreflective desire.

III. The Value of Authenticity

Again, in comparing these models of authenticity, my aim is to consider which one presents a preferable concept of authenticity. The concept of authenticity,

18. Rogers, *On Becoming a Person*, 341. Rogers notes, as a corollary to congruence, that "If an individual is at this moment entirely congruent, his actual physiological experience being accurately represented in his awareness, and his communication being accurately congruent with his awareness, then his communication could never contain an expression of an external fact. . . . If the person is thoroughly congruent then it is clear that all of his communication would necessarily be put in a context of personal perception."

as I argued in Section I, serves an interpretive function, making meaning of certain actions as "true to oneself." In light of a subjectively constituted understanding of the self, we can see that a concept of authenticity is to some extent prescriptive—what behavior we interpret as "authentic" can shape how we behave, as we seem to want to "be ourselves." Considering both the spontaneity and congruence models to be adequately plausible as accounts of what authenticity is, the preferable model will be the one that constructs our "real selves" toward the best results. So, what kind of behavior does each model encourage in us?

First, let's look back at our illustrations. The spontaneity model, in that it considers the unreflective self to be authentic, says that I am only being myself if I do not adjust my conduct in the moment based on how I expect it to be received by my father. That seems to amount to a sort of "flying blind": authenticity is whatever comes out of my mouth in the heat of the moment. As my spontaneous "real self" is still socially influenced, I will be more likely to follow my cultural conditioning regardless of its consequences. While I can hope that I have been conditioned to spontaneously act well, various socializations will be wrapped up in that conditioning: I may display anger rather than vulnerability based on gendered socializations, for example, or lash out physically if that is how I have seen anger expressed. If I act spontaneously, it seems that both situations would be likely to play out worse, straining our relationship and placing raw feelings in the way of interpersonal understanding. If I act congruently, on the other hand, I voice my dissatisfaction and anger in ways that my father will be more likely to understand. Self-awareness and self-reflection are valuable tools which make me act more ethically as well as more strategically; in placing those tools necessarily outside of my "real self," the spontaneity model discourages me deepening them as habits and incorporating them into my sense of self. The congruence model, on the other hand, as it allows me to understand deeply reflective persons as authentic, makes space for a sense of my "real self" that is more socially connected and attuned.

I do not want to rely entirely on my illustrations to make my case, however. For a second sense of the relative merits of these conceptions, let's look to the value that our thinkers themselves place on authenticity.

Larmore, who declares that authenticity, "being essentially unreflective, cannot be the goal of a project,"[19] still ascribes a heavy value to personal authenticity. Larmore writes that "the ultimate insight in Stendhal's conception of authenticity . . . is the recognition of the error in the notion of a *life-plan*; . . . the importance of the Romantic theme of authenticity is that it disabuses us of the idea that life is necessarily better the more we think about it."[20] Larmore does not dwell on the implicit middle premise of his argument, but we can tease it

19. Larmore, *The Romantic Legacy*, 93.
20. Larmore, *The Romantic Legacy*, 93.

out: to get from the claim that authenticity is unreflective to Larmore's sense that that life is better if unreflective, the necessary middle step is that *life is better if authentic*. Larmore, maintaining that authenticity must be without reasons (recall that reasons, according to Larmore, necessarily consider others' judgments), writes that, "the urge to 'be ourselves,' however, does not arise from our commitment to being rational, . . . we ought not to suppose that there is anything more to be said."[21] This seems, frankly, rather evasive. One is left with the feeling that Larmore values personal authenticity as more or less an end in itself, which seems like a pre-judged preference toward a certain kind of "independent" self.

Mitchell identifies authenticity as a component of psychological health, claiming that authenticity simply feels *better* than inauthenticity: an authentic act or self-expression "feels 'right,' suits both the external, interpersonal context and the internal emotional context," where an inauthentic one "feels 'off,' forced, contrived, out of 'sync' interpersonally, internally, or both."[22] That "right" feeling is the "subjective sense of greater depth,"[23] "a kind of freedom and authenticity that is both rare and precious."[24] Mitchell's position seems like a weaker phrasing of Larmore's thought process—too much social interference simply feels *wrong*.

Rogers, on the other hand, sees authenticity as a means for strengthening interpersonal connection and becoming a "fully functioning person."[25] As a move away from defensiveness rather than social influence as such, according to Rogers, authenticity allows a person to be more open to experience, more fluidly adaptable, more trusting of their intuition, and more sensitive to and accepting of a full range of feeling; in sum, to experience a greater richness of life.[26] In the claim about the richness of life, that is fairly aligned with Mitchell—the sense of alienation or forced-ness that comes with inauthenticity as non-spontaneity may also interfere with openness to experience and sensitivity to feelings. Rogers adds, however, that authenticity as congruence greatly improves interpersonal relationships. Rogers's articulation of the concept of congruence appears in his essay "A Tentative Formulation of a General Law of Interpersonal Relationships," in which he claims that greater levels of congruence in any relationship will lead to "a tendency toward reciprocal communication with a quality of increasing congruence; a tendency toward more mutually accurate understanding of the communications; improved psychological adjustment and functioning in both parties; mutual satisfaction in the relationship."[27] These are all grand claims, which Rogers claims to have drawn from his psychotherapeutic research

21. Larmore, *The Romantic Legacy*, 96-97.
22. Mitchell, *Hope and Death*, 131.
23. Mitchell, *Hope and Death*, 140.
24. Mitchell, *Hope and Death*, 150.
25. Rogers, *On Becoming a Person*, 183.
26. Rogers, *On Becoming a Person*, 187-195.
27. Rogers, *On Becoming a Person*, 344.

and observation of clients, for the possibilities that personal authenticity can open up. The point being that in the congruence model, rather than simply being a good feeling in itself, authenticity becomes a useful means toward ends both personal and interpersonal, both of which point toward better living.

IV. Conclusion

When we use the concept of authenticity, we are performing an interpretive act, drawing a meaningful line between what is and what is not a real expression of the self. To draw that line to include more self-reflection seems closer to the common usage of being one's "truest self," as well as providing a reason to consider authenticity as a virtue, which we already tend to do. I do not claim to have presented an airtight case for a congruence view of authenticity over a spontaneity view. As I have said, they both seem like plausible ways of interpreting behavior as self-expressive. Considering its consequences, however, congruence seems preferable. Understanding authenticity under a congruence model would help to construct selves that are more ethical, socially attuned, "fully functioning" in Rogers's sense of general well-being—where operating under the spontaneity model encourages a relatively isolated, reactive, un-self-aware mode of being. Under the congruence model of authenticity, far more than the spontaneity model, "be yourself" actually seems like good advice.

Bibliography

"Authenticity." *OED Online.* Oxford University Press, December 2018. www.oed.com/view/Entry/13325.

Butler, Judith. "Imitation and Gender Insubordination." In *Women, Knowledge, and Reality*, edited by Ann Garry and Marilyn Pearsall, 371-387. New York: Routledge, 1996.

Larmore, Charles. *The Romantic Legacy.* New York: Columbia University Press, 1996.

Mitchell, Stephen A. *Hope and Death in Psychoanalysis.* New York: Basic Books, 1993.

Rogers, Carl. *On Becoming a Person: A Therapist's View of Psychotherapy.* Boston: Houghton Mifflin, 1961.

Richard Swinburne's False Dilemma

Owen Crocker
University of the Fraser Valley

Abstract

Richard Swinburne recently released a paper titled, "Causation, Time and God's Omniscience." In this paper, Swinburne argued that God's omniscience must be understood in a way that excludes divine foreknowledge. Swinburne deems this a necessary step in order to protect our freedom of the will. The purpose of my paper will be to refute Swinburne's central argument. The goal of refuting Swinburne's argument is to maintain the possibility of the compatibility of both divine foreknowledge and free human agency.

Introduction

Richard Swinburne recently released a paper titled, "Causation, Time and God's Omniscience." In this paper, Swinburne argues that God's omniscience must be understood in a way that excludes divine foreknowledge of human actions. Swinburne deems this a necessary step in order to protect our freedom of the will. Swinburne claims that if God possesses foreknowledge of our future actions and humans possess free will, it is possible to prove God's foreknowledge wrong. Therefore, because God's knowledge cannot be wrong, God must not possess knowledge of our future actions. The purpose of this paper will be to refute this central argument that Swinburne presents. By refuting Swinburne's argument, I will be defending the compatibility of God's divine foreknowledge of human actions and human freedom.

What Is at Stake? Scripture, Sovereignty, and Salvation

The purpose of this section is to explain the theological importance of claiming, as Swinburne does, that God does not know our future actions. Clive S. Lewis writes, "Everyone who believes in God at all believes that He knows what you and I are going to do tomorrow."[1] This quote captures how the vast majority

1. Clive S. Lewis, *Mere Christianity* (New York, NY: Macmillan, 1952), 148.

of Christians, including myself, understand God's divine omniscience regarding our future actions.

It is important to note that God's knowledge is considerably different than human knowledge. Human knowledge is often mistaken. However, God's knowledge is infallible, meaning, that God cannot be mistaken. William L. Craig and James P. Moreland write, "God is omniscient if God knows every true proposition."[2] To clarify, a proposition can be understood as the content of a statement. For example, a propositional statement might say, "The Canucks will win the Stanley Cup in 2022" or "Today is June 7th, 2001." These statements are either true or false, based on the content of the statement. As it was stated previously, God's knowledge is always correct and never mistaken. So, when God's knowledge is mentioned, it should be understood as infallible knowledge of all past, present and future propositional statements, including our future actions.

However, if God does not know our future actions, there are three major theological themes that every Christian will need to reconsider. The first of these areas is *scripture*. Traditionally, Christianity has held that God exhaustively knows the future, including our actions. This belief has primarily been built upon the Christian scriptures. As an example, Matthew 26:31-35 tells a story of Jesus predicting that his disciple, Peter, will publicly deny knowing him three times. Later on that night, Peter denies Jesus three times. This story appears to demonstrate that God knew Peter's future actions. In addition to this story, there are many other biblical passages that give warrant to the view that God knows the future exhaustively.[3] If one accepts that God does not know our future actions, many verses that support the traditional conception of God's foreknowledge would need to be reinterpreted.

The second area that would require reconsideration is the Christian idea of God as *sovereign*. There is a general consensus among Christian believers that God is in control, and that one should not worry about the future. Proverbs 16:9 states, "The heart of man plans his way, but the Lord establishes his steps." This verse indicates that God has a plan and purpose for our lives, as he has already established our steps. The idea that God has made a purpose for all the things that will happen to us is extremely comforting to many Christians. That being said, the idea that God is sovereign over our lives becomes rather puzzling when one no longer holds that God knows our future actions. One might ask: How could God have a plan and purpose for my life if he does not know what will happen?

2. William L. Craig and James P. Moreland, *Philosophical Foundations for a Christian Worldview*, 2nd ed. (Downers Grove, Il: InterVarsity Press, 2017), 524.

3. In the book, Bruce Ware, *God's Lesser Glory: The Diminished God of Open Theism* (Wheaton, Il: Crossway, 2000), Bruce Ware mentions other verses such as, Isaiah 40-48, Psalm 139, Daniel 11 and John 12-13. These verses would also need to be reinterpreted if one denies God's foreknowledge of human actions.

Lastly, there are questions that arise regarding salvation. As John Piper writes, "Open theism's[4] denial of God's exhaustive definitive foreknowledge renders uncertain the execution of God's plan of salvation through the delivering up of his son by crucifixion on the cross."[5] The arc of the biblical narrative points to the fact that God had a pre-orchestrated plan regarding how he would make salvation possible for all people. This plan involved a number of people who were involved in Jesus' crucifixion. These people include: Herod, Pontius Pilate, Gentiles, soldiers and the peoples of Israel (Acts 4:26-28). If God did not know that these agents would carry out the crucifixion of Jesus, God's plan of salvation might not have come to fruition, and Jesus may not have been crucified.[6] This is a conclusion that many Christians will be hesitant to accept. In this sense, God's plan of *salvation* is another area that Christians would need to reconsider, if one were to deny God's foreknowledge of human actions.

The three theological themes of scripture, sovereignty and salvation are some of many possible Christian beliefs that would need to be reconsidered if God does not know our future actions. As Sandra Visser writes, "This option is . . . something I think one ought not to take lightly."[7] Therefore, the Christian has much at stake regarding God's foreknowledge of human actions as it will inevitably impact many spiritual beliefs as well as personal convictions.

Divine Omniscience and Freedom of the Will

Before I begin to evaluate Swinburne's argument, it is important that the general problem of theological fatalism is properly understood. Anselm of Canterbury writes, "It certainly seems as though divine foreknowledge is incompatible with there being human free choice. For what God foreknows shall necessarily come to be in the future, while the things brought about by free choice do not issue from necessity. And if divine foreknowledge and human free choice cannot both exist, it is impossible for God's foreknowledge, which foresees all things, to coexist with something happening by free choice."[8] What Anselm is essentially saying is this: How can I have freedom in my future actions, if God knows what those future actions will be? If God knows what I am going to do tomorrow, am I free to do as I please? Within these questions, there is a tension between two fundamental principles. These fundamental principles are:

(1) God contains knowledge of future human actions.

4. Open theism is the theological term for the view that God does not know our future actions.

5. John Piper, Justin Taylor, and Paul K. Helseth, *Open Theism and the Undermining of Biblical Christianity* (Wheaton, Il: CrossWay, 2003), 320.

6. If the reader is unfamiliar with the story of Jesus' crucifixion, I would encourage them to read Matthew 27:1-54, Mark 15:1-40, Luke 23:1-48 or John 19:1-30.

7. Sandra Visser, "God's Knowledge of an Unreal Future," in *Philosophical Essays Against Open Theism*, ed. Benjamin H. Arbour (New York, NY: Routledge, 2019), 40.

8. Anselm of Canterbury, *The Major Work*, edited by Brian Davies and G. R. Evans (Oxford: Oxford University Press, 2008), 434.

(2) Humans possess freedom of the will.

As one can see, there appears to be a contradiction between these two principles. This apparent contradiction, as mentioned before, is the problem of theological fatalism. Succinctly stated, theological fatalism is, "the thesis that infallible foreknowledge of a human act makes the act necessary and hence unfree."[9] Richard Swinburne acknowledges this problem and proposes a solution that entails that we reconsider the first principle. We will now turn to Swinburne's argument.

Swinburne's False Dilemma: Part One (God's Timelessness)

Swinburne has composed his paper in two major sections. These sections are (1) God as timeless, and (2) God as temporal. When one speaks of God as timeless, one typically means that God is beyond time, meaning, and that "he exists but does not exist at any point in time and he does not experience temporal succession."[10] In contrast, when God is referred to as being temporal, this means that God experiences temporal succession. In other words, God experiences some events before other events. An analogy I find helpful to distinguish between timelessness and temporality is to imagine an author and a book.[11] If God is timeless, imagine God being an author of a book. As the author, God sits outside the pages of the book and can see all the pages and contents at the same time. In reality, if God was timeless, he would sit outside our current timeline and see all events at once, without temporal succession. However, if God is temporal, his experience would be similar to the characters in the book, as they are limited to experiencing the book's contents in succession. The temporal characters experience page one before two and two before three and so on. In reality, if God was temporal, he would experience Monday before Tuesday, and Tuesday before Wednesday and so on.

Given that there are two major conceptions of God's relationship with time, Swinburne addresses both those who hold that God is timeless as well as those who hold that God is temporal. Swinburne proactively addresses a problem he might encounter if he were only to address one of the conceptions of God's relationship with time. For example, if Swinburne claimed that God could not know the future because he is temporal, one could easily refute Swinburne by simply claiming that God is timeless. However, Swinburne addresses both God as timeless and temporal in his paper, thereby, forcing everyone to respond to his argument.

9. Linda Zagzebski, "Foreknowledge and Free Will," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (2017), URL = <https://plato.stanford.edu/archives/sum2017/entries/free-will-foreknowledge/>.

10. Gregory E. Ganssle, "God and Time," in *Internet Encyclopedia of Philosophy*, https://www.iep.utm.edu/god-time/

11. This analogy is used in Lewis, *Mere Christianity*, 147.

As many have argued, if God is outside of time, then God's knowledge is incapable of necessitating our actions. This particular response to theological fatalism is titled, *the Boethian solution*. Boethius' understanding of God being outside of time allowed him to claim that God's knowledge of events that are future to us, are actually present to God. Therefore, God's knowledge of our future events does not necessitate that we choose the action that he knows we will choose. Boethius writes, "So if you should wish to consider his foreknowledge, by which he discerns all things, you will more rightly judge it to be *not foreknowledge as it were of the future but knowledge of a never-passing instant* . . . And therefore this divine foreknowledge does not alter the proper nature of things, but sees them present to him, just such as in time they will at some future point come to be."[12]

Swinburne does not directly respond to this argument. Rather, he responds to the general conception of God being outside of time. Swinburne writes, "I conclude that it seems almost impossible to give any sense to the view that there could be a timeless God who knows what is happening in the physical universe and causes events in that universe."[13] Swinburne's argument claims that if God is timeless, and the world is in time, it is a contradiction to claim that God can be causally active in the world. The idea that God could cause an event in a temporal world entails that God is related to the world. But, if God is related to the world, then it appears he is in time. This argument can be summarized as follows:

> *Premise 1*: God is creatively active in the temporal world.
>
> *Premise 2*: If God is creatively active in the temporal world, God is really related to the temporal world.
>
> *Premise 3*: If God is really related to the temporal world, God is temporal.
>
> *Conclusion*: God is temporal.[14]

One could deny Premise 1 and hold that God is not creatively active in the temporal world. But, if one would like to maintain the idea that God is a *personal* God (as is the God in Christianity), then that is not a viable option. I am inclined to accept this argument, and agree with Swinburne that it does not seem rational to hold that a timeless God could have any causality in the temporal world. Therefore, if one is unable to refute the above argument, you must come to the conclusion that God is temporal (or, in other words, exists in

12. Boethius, *The Theological Tractates: The Consolation of Philosophy*, edited by Hugh F. Stewart and Edward K. Rand (London: Heinemann, 1918), 427. Emphasis added.

13. Richard Swinburne, "Causation, Time, and God's Omniscience," *Topoi: An International Review of Philosophy* 36, no. 4 (2017): 683.

14. This argument was formulated in William Lane Craig, *Time and Eternity: Exploring God's Relationship to Time* (Wheaton, Il: Crossway, 2001), 87.

time).

Swinburne's False Dilemma: Part Two (God's Temporality)

Swinburne's next step in his argument is to show why a temporal God is unable to know our future free actions. However, this time he is not going to attack the general idea of a temporal God (as he did with a timeless God), but the idea that a temporal God knows our future actions. Swinburne's argument can be summarized as follows:

Premise 1: God knows at T^1 that Peter will deny Jesus at T^2.

Premise 2: God's knowledge entails a belief of what Peter will do at T^2 as well as the truth of the matter of what Peter will do at T^2.

Premise 3: At T^2 it is within Peter's freedom to deny or not deny Jesus.

Premise 4: Based on *Premises 1-3*, Peter is able to prove God's knowledge at T1 to be false.

Premise 5: *Premise 4* is a logical contradiction.

Conclusion: Therefore, one of the premises must be wrong.

As stated in *Premise 5*, there is a logical contradiction based on the fact that Peter is able to prove God's knowledge wrong. If we apply the principle of the law of non-contradiction to the question at hand, it is nonsensical to say Peter will deny Jesus and Peter will not deny Jesus. The law of non-contradiction states that, something cannot at the same time be and not be. Therefore, these two statements cannot both be true in reference to the same event. God cannot hold a true belief that entails the truth that Peter will deny Jesus at T^2 at the same time as Peter not denying Jesus at T^2. These two statements are incompatible with one another.

Swinburne rightly recognizes this problem. In response to the contradiction, Swinburne concludes that *Premise 1* is the incorrect premise. Swinburne writes, "God's omniscience must be construed in the *weaker sense*."[15] The term "weaker sense," to Swinburne, means the elimination of the God's knowledge of future human actions. However, God's weaker sense of divine omniscience still includes all other propositional statements. So, as long as a propositional statement does not rely on the decision of a future human action, which could change the truth-value of the statement, God's knowledge would include that proposition. For example, the statement, "Peter will deny Jesus three times" clearly rests on Peter's future action to deny Jesus or not to deny Jesus. Therefore, God would not be able to know the truth of this proposition. But, the statement, "There

15. Swinburne, *Causation*, 683. Emphasis Added.

are currently 30,000 red cars driving in Vancouver" does not seem to rely on future human actions. So, God's knowledge would include that proposition.

Swinburne solves the problem of theological fatalism by understanding God's omniscience in the "weaker sense." In Swinburne's mind, he has now shown that God is unable to know the future regardless of whether one holds that God is timeless or temporal. However, as stated in the prior section titled "What Is at Stake?", there are theological reasons for the Christian to hesitate before accepting Swinburne's solution, especially when there are other possibilities available.

Responding to Swinburne's Dilemma: The Middle Road

My response to Swinburne is twofold. First, I will address the dilemma of God's relationship with time. Secondly, I will address which premise must be revised in order to resolve Swinburne's argument against a temporal God knowing our future actions. My response will also defend both principles that the problem of theological fatalism is premised on.

Swinburne presented us with the dilemma of God either being temporal or timeless. However, in my opinion, this is a false dilemma. It is a false dilemma because there is a third option, a middle road. There are many philosophers starting to accept the idea of God being both timeless and temporal. I am not suggesting that God is simultaneously timeless and temporal. Rather, I am suggesting that God *was* timeless and is *now* temporal.

In respect to which conception of God's relationship with time one adopts, they all appear to be equally valid assumptions. However, although they are all equally valid assumptions, one must carefully consider what follows from the conception one holds. For example, if one holds that God is timeless, it seems that God cannot be causally active in the world. That being said, I will assume the conception that God *was* timeless and is *now* temporal, and then demonstrate how this conception resolves the problem of theological fatalism.

William Lane Craig writes, "The most plausible view of God's relationship with time is that He is timeless without creation and temporal subsequent to creation."[16] However, simply appealing to God's changing relationship to time does not adequately address all of Swinburne's claims. Moving forward, I will use Craig's understanding of God's relationship with time as a framework to show the compatibility of God's knowledge of future human actions and human freedom.

The Middle Road: Part One (God's Timelessness)

As I suggested, the middle road entails that God was timeless before creation

16. Craig, *Time and Eternity*, 236.

and temporal subsequent to creation. If we consider God's knowledge of human decisions before creation, there is no implication of the necessity of those future decisions. What I mean is that the problem of theological fatalism has yet to arise, because there are no humans to apply the problem to.

In God's timeless state, it is plausible[16] to assume that God has knowledge of what would happen if he were to create a world (or possible worlds). Possible worlds refer to alternative worlds that are not in fact the case. For example, one could imagine a possible world where Stephen Harper won the 2015 Canadian election, rather than Justin Trudeau. However, in the actual world, Justin Trudeau won the 2015 Canadian election. The knowledge of possible worlds that God possesses would be based on what would happen if specific events were to be actualized. For example, in 2011 the Vancouver Canucks lost to the Boston Bruins in the Stanley Cup finals. However, in a possible world, one can imagine two different teams playing in the 2011 Stanley Cup finals. I am suggesting that God knows the outcome of all the possibilities of the 2011 Stanley Cup finals, regardless of the two teams playing one another. We can apply this knowledge of future hypothetical events to all possibilities of what could happen in the world, giving God knowledge of all possible worlds that he could have created. God's knowledge of possible worlds would also include what humans would do in various circumstances, if those circumstances were actualized.

One should take note that Swinburne's argument is unable to be used against a timeless God as long as there is no creation. This is because Swinburne's main contention is that a timeless God cannot be creatively active in the world. As a reminder, Swinburne argues that if a timeless God is creatively active in the world, then God is really related to the world. But, if God is really related to the world, then God is temporal. Therefore, a timeless God cannot be creatively active in the world, because that would deem God temporal. But, without a world to be creatively active with, the argument is irrelevant. Therefore, it seems that a timeless God could have knowledge of future human actions without necessitating what will come to pass. However, we have yet to address what happens subsequent to creation.

The Middle Road: Part Two (God's Temporality)

The difficulty appears after the initial time boundary of creation. Now that God has created a certain world, he possesses knowledge of the actual world. His knowledge of what would happen if he created a possible world has now become foreknowledge of what will happen. In other words, God had knowledge of all possible worlds in his timeless state. But, after creation, God also has knowledge of the actual world, including what will come to pass. We now need to respond

16. There are some who would not consider this knowledge plausible, which will be addressed in the section titled "Objections Part Two: The Grounding Objection."

to Swinburne's argument against a temporal God being able to know future human actions.

Before I continue, there is a distinction that needs to be made. The distinction is between chronological and logical priority. These two concepts often get confused when discussing the implications of God's foreknowledge of our future actions. For example, it is easy to assume that since God has knowledge of what will happen at T^1, his knowledge necessitates the action at T^2, which negates our freedom. However, the fact that God's knowledge is chronologically prior to our actions on a given timeline, does not entail that God's knowledge is logically prior to our actions. The truth of what humans would do in a given scenario is logically prior to God knowing what we will do in that scenario. For example, if Bob were to propose to Sally, and if Sally were to say yes, God's foreknowledge would include this. But, if Sally were to say no to Bob, then God's foreknowledge would include Sally's dismissal of the proposal. Either way, the truth of what one does in a given circumstance is the logical basis for God's foreknowledge. When God creates the world, his knowledge that was based on the logical priority of our actions will become chronologically prior to when we will carry out those actions. But, the chronological priority in God's temporal state, is founded on the logical priority of the truth of our actions. In other words, the truth of what we would do in certain circumstances determines God's knowledge, not the other way around. This distinction between logical priority and chronological priority demonstrates that God's chronologically prior knowledge of our future actions does not imply that our future actions are not free decisions.

To restate, in God's timeless state his knowledge is based on the truth of what would happen if he created a certain world. Therefore, since God's foreknowledge is based on our future acts, his knowledge does not necessitate what we do. Rather, what we will do necessitates God's foreknowledge, as we saw with the example of Sally's response to Bob's proposal. That being said, there is a sense in which one will act in accordance to what God knows we will do (regardless of why God knows what we will do). For example, since God knows at T^1 that Peter will deny Jesus at T^2, Peter will deny Jesus at T^2. However, the fact that Peter will do what God knows Peter will do, does not negate his freedom.

Are We Free? The Principle of Alternate Possibilities

Now that I have addressed Swinburne's false dilemma of God's relationship with time, I will now address which premise must be revised in order to sufficiently answer Swinburne's argument. In contrast to Swinburne rejecting *Premise 1*, which states, "God knows at T1 that Peter will deny Jesus at T^2," I suggest that *Premise 3* is the incorrect premise. *Premise 3* states, "At T^2 it is within Peter's freedom to deny or not deny Jesus." Rejecting *Premise 3* may appear to be at odds with the second principle of theological fatalism, which states, "Humans possess freedom of the will." However, I propose that one can have

free will without having the ability to do otherwise than what one did. In order to understand my proposition, we must consider the *principle of alternate possibilities* (PAP).[17]

> PAP: A person is morally responsible for what she has done only if she could have done otherwise.[18]

The principle of alternate possibilities is first and foremost concerned with moral responsibility. However, the principle can also be useful for understanding the concept of freedom of the will. For example, I will adjust PAP into a form that is concerned with freedom of the will. I will call this version of the principle of alternate possibilities PAP2 to avoid confusion.

> PAP2: A person possesses freedom of the will only if she could have done otherwise.

Most people will be inclined to accept both PAP and PAP2 on a prima facie basis. However, I am going to argue that upon further consideration, both PAP and PAP2 are highly questionable. PAP2 states that one possesses freedom of the will only if one could have done otherwise. Another way of putting the principle is to say that freedom of the will requires the ability to do the opposite of what one chooses. When we consider PAP2, it is often thought about in past events. For example, let's consider Peter's denial of Jesus. We are inclined to agree that if Peter could not have done otherwise than what he did, he was not free. However, if we consider future actions, the principle starts to become questionable.

For my example, I will use the loss of the Canucks to the Bruins in the 2011 Stanley Cup finals. After the game, I am approached by a few Bruins fans and am asked, if I am a Canucks fan. It appears I have two options. I can say yes, and risk the possibility of humiliation or I can say no and lose my integrity. However, there is an additional twist. When I was born, my father implanted a chip into my brain that forces me to say yes, whenever I am asked if I am a Canucks fan and I am about to say no to the question. So, if I were ever to say no to the question of whether I am a Canucks fan or not, the chip in my brain would kick in and force me to say yes. According to PAP2, it seems that I am not free to answer the question. However, what happens if I were to say yes voluntarily, under my own free choice? The implanted chip did not influence my decision at all. I chose by my own free will to risk the humiliation and say, "Yes, I am a Canucks fan." This sort of example is often used to show why one's freedom does not require the ability to do otherwise. Therefore, PAP2 seems to

17. The *Principle of Alternative Possibilities* was originally coined by Harry Frankfurt, "Alternative Possibilities and Moral Responsibility," *The Journal of Philosophy* 66, no. 23 (1969): 829-839.

18. This phrasing of PAP was based on Peter van Inwagen, *Thinking About Free Will* (Cambridge, UK: Cambridge University Press, 2017).

be highly questionable.

Divine Omniscience and The Principle of Alternate Possibilities

The typical Frankfurtian examples do seem to have a fatal flaw. This flaw being, the causal mechanism (CM) that kicks in if you were to do the opposite of whatever it was programmed to prevent you from doing. The CM appears to limit our freedom, as it is possible that the CM kicks in and causally forces us to go against our free will. However, since we are discussing divine omniscience, it would be helpful to replace the CM with God's foreknowledge (GF) as the mechanism that could potentially disable us from doing what we will in fact do. There is a significant difference between GF and CM. This difference is rooted in causality. CM entails a physical cause that forces us to go against our will, which is a problem. However, God's foreknowledge does not entail any physical causation, but is simply knowledge of what we will do. In other words, God's foreknowledge has no causal implications whatsoever. As stated previously, the outcome of our actions is what God's foreknowledge is based on and is determined by. In other words, there will never be a possibility for our actions to be impacted by GF because our actions will always align with God's foreknowledge. Again, this alignment is due to God's knowledge being based on what we will do in the future (logical priority).

That being said, it is common for those who are engaging in this material for the first time to ask: Why must I still do what God knows I will do, even if his knowledge doesn't necessitate my action? The answer is quite simple really, that being, the question is mistaken. It is not a matter of what Sally must or must not do. It is a matter of what Sally will do freely. If it is true that Sally will say yes to Bob's proposal, then God's foreknowledge would account for that and vice versa. Sally will say yes to Bob's proposal not because God's foreknowledge entails that she will, but because that is what Sally will freely choose to do in that scenario. Thereby, as stated previously in this section, God's foreknowledge has no causal implications because it is rooted in the logical priority of what humans would do given various circumstances.

Freedom of the Will: What Does It Require?

However, one might still be hesitant to agree that one possesses freedom of the will even if they have the ability to choose freely and cannot be causally impacted by God's foreknowledge. One might claim that real freedom needs to be accompanied by the genuine option of alternate possibilities. In response, I argue that genuine freedom requires four components. These four components were originally argued for by St. Augustine, and later reformulated in David Hunt's paper, "On Augustine's Way Out" (1999). These four components include:

Possession: The act/decision belongs to the agent/person alone.

Approval: The agent approves or looks favourably upon the decision/act that they are performing.

Power: The causal power to perform the act belongs to the agent/person.

Lack of Compulsion: The agent's action is not necessitated by any coercive (or compulsive) causal force.

If one's decision at moment T is accompanied by all four of these components, it seems difficult to claim that the person did not make a free decision. That being said, adding the extra condition that this person must have had the ability to do otherwise seems irrelevant. It seems to me that these four components are sufficient for freedom. Those who claim that one needs the ability to do something they would not choose to do, seem to be desiring a sort of hyper-freedom of the will (HFW). HFW requires not only that one has possession of their decision, approval of their decision, power of their decision and lack of any compulsory forces but, also that they can do what they would not do in a given situation. It seems to me that this sort of free will is more than we normally deem necessary. Why does one need the ability to do something that they would not choose to do in a given scenario? It seems to me that both PAP and PAP2 can be refuted by demonstrating that freedom of the will can be met by the previously listed four components.

The End of the Road: Freedom and Foreknowledge

It seems to me that I have sufficiently answered Swinburne's argument, while defending the compatibility of God's foreknowledge of future human actions and human freedom. To restate my general argument, God in his timeless state possessed knowledge of our future actions based on the truth of what we would do in freedom-permitting circumstances. After the act of creation, God retained this knowledge as he entered into his temporal state. Despite God knowing what we will do in the future, we still possess freedom of the will for two primary reasons. First, his foreknowledge is based on how we would act; therefore, he is not determining our actions. Secondly, our freedom does not require alternate possibilities. Despite this solution solving many problems, there are still possible objections. The last portion of my paper will address two possible objections.

Objections Part One: Moral Responsibility

The first objection is in reference to the denial of the principle of alternate possibilities. In previous sections, I have argued that freedom of the will does not require the ability to do otherwise. However, traditionally, PAP is concerned with moral responsibility. Freedom of the will and moral responsibility are separate philosophical concepts that often get conflated. So, it is important to address them separately to avoid confusion. This portion of the paper will address moral responsibility and whether or not it requires the ability to do

otherwise. I will argue that it does not, and that humans can be held morally responsible without the ability to do otherwise.

There have been additional principles that have developed in order to weaken Frankfurtian-type examples. As I reminder, Frankfurtian-type examples are those which intend to demonstrate the reasonableness of denying PAP. In this paper, I used the example of a father implanting a chip into the son's brain regarding the Canucks, which can be found on pages fourteen and fifteen. These additional principles include PPA, PPP1 PPP2 (Van Inwagen 2011, 245):

> PPA: A person is morally responsible for failing to perform an act only if he could not have performed that act.
>
> PPP_1: A person is morally responsible for a certain event-particular only if he could have prevented it.
>
> PPP_2: A person is morally responsible for a certain state of affairs only if (that state of affairs obtains and) he could have prevented it from obtaining.[19]

These three principles are all very similar due to the fact that the requirement for moral responsibility is 'the ability to prevent your decision'. So, one could argue that if God knows Peter will deny Jesus, Peter cannot be held morally responsible if he could not have prevented his denying of Jesus from actualizing. As other philosophers have pointed out, there is an underlying more fundamental principle in all these other principles. As Immanuel Kant states:

> If it is morally obligatory for one to do something, then one can do it; and if it is morally obligatory for one to refrain from doing something, then one can refrain from doing it.[20]

This principle is attempting to formulate the very basis for what moral responsibility consists of. According to Kant and others, in order to be morally responsible, one must be able to actualize the moral obligation or prevent the morally reprehensible act from occurring. For example: If one should not lie, then one must be able to prevent oneself from lying. However, based on what meta-theory one uses to for their grounds of morality, one may come to a different conclusion. Michael McKenna denies this fundamental principle of moral responsibility and suggests another. McKenna writes, "A persons' moral responsibility concerns what she does do and her basis for doing it, not what else she could have done."[21]

19. Van Inwagen, *Thinking About Free Will*, 245.

20. Quoted in John Martin Fischer, "Frankfurt-Type Examples and Semi-Compatibilism," in *The Oxford Handbook of Free Will* (2nd. ed.), ed. Robert Kane (New York, NY: Oxford University Press, 2011), 293.

21. Michel McKenna and Derek Pereboom, *Free Will: A Contemporary Introduction* (New

I am inclined to accept Mckenna's principle (titled as *the L-Reply*) over Kant's because I am not sure that the possibility to do otherwise is needed in what we deem to be a free and morally responsible event. As we saw earlier, it seems to me that if one possesses the four components of possession, power, approval and lack of compulsion, then one is making a free decision. This free decision seems to align with McKenna's principle of what moral responsibility consists of. The four components will render the person not only free to act but also the ability to have reasons for the act itself. Therefore, the four components mentioned earlier (i.e., approval, possession, power and lack of coercion) are sufficient to allow the person to meet the criteria of Mckenna's *L-Reply*. In sum, if one were to raise the objection of moral responsibility, I think this objection fails as there is an adequate account of how one can be held morally responsible without the ability to do otherwise.

Objections Part Two: The Grounding Objection

As stated in my previous section titled "The Middle Road: Part One (God's Timelessness)", I make an assumption that the careful reader may have picked up on. In that section I stated, "In God's timeless state, it is plausible to assume that God has knowledge of what would happen if he were to create a world (or possible worlds)." This sort of knowledge contains what individuals would do in hypothetical scenarios, or counter-factual states of affairs. This sort of knowledge has come to be known as middle knowledge (MK). For the most part, theologians and philosophers agree that God has natural knowledge (NK) and free knowledge (FK). But, middle knowledge is not as widely accepted. These three categories of God's knowledge can be summarized as follows:

> NK: God knows the range of possible worlds (everything that could happen).
>
> MK: God knows the rang of feasible worlds (everything that would happen).
>
> FK: God knows the actual world (everything that will happen, after creation).

Those who are skeptical of middle knowledge base their skepticism on what is called *the grounding objection*. Steven B. Cowan writes, "The basic idea behind the grounding objection is the contention that God cannot have middle knowledge because the counterfactuals of freedom which are the objects of His middle knowledge have no truth-value. That is, there are no actual state of affairs to which such propositions correspond to in order to provide a truth condition for their truth or falsity."[22] According to Cowan, along with many others, counterfactual states of affairs do not have a truth-value, and therefore, God cannot know the truth of the statement. William Hasker, another opponent

York, NY: Routledge, 2016), 104.

22. Steven B. Cowan, "The Grounding Objection to Middle Knowledge Revisited," *Religious Studies* 39, no. 1 (2003): 93-102.

of Molinism,[23] states, "In order for a (contingent) conditional state of affairs to obtain, its obtaining must be grounded in some categorical state of affairs. More colloquially, truths about 'what would be the case . . . if' must be grounded in truths about what is in fact the case."[24] As you can see in both quotes, Cowan and Hasker are essentially saying that propositions must have a corresponding state of affairs to prove that proposition true or false. These claims are premised on what is known as truth-maker theory, which bears resemblance to the logical positivist movement. Truth-maker theory requires a state of affairs in the world to make a statement true. For example, the statement "it is snowing outside" would in fact be true, if it was snowing outside. The fact that there is snow falling is the truth maker for the statement, thereby, proving it to be true. However, a statement like, "If Q were in situation X, Q would do Y," does not have a corresponding truth maker to prove the statement true or false.

The Molonist (proponent and defender of middle knowledge) could begin their response to the grounding objection by stating that counterfactuals are for the most part assumed to be true or false for many reasons. First, we often use counterfactual statements in our everyday affairs, which presupposes that counterfactuals do have a truth-value. Second, the law of the excluded middle seems to indicate that many counterfactuals must be either true or false. For example, take the two following statement, "$(P \supset Q)$ or $(P \supset \neg Q)$," only one of these can be correct, seeming to indicate that counterfactuals do have a truth-value. Thirdly, there is also scriptural evidence[25] to support that God contains middle knowledge.[26] However, these three reasons are simply to support the idea that counterfactuals are a reasonable position to hold, and thereby seem to indicate that those who claim counterfactuals don't have a truth maker are the ones who carry the burden of proof.

However, there is even a greater reason to doubt the grounding objection, that being the insufficiency of truth-maker theory. For example, take the statement "Dinosaurs are extinct today." This negative statement does not seem to have a truth-maker, yet almost anyone would agree that it is true. But, how can it be true if it does not have a truth maker? Or take the statement, "The next Canadian Prime Minister will be a woman." This seems to either be true or false, but there is no current truth-maker or state of affairs that can prove this to be true or false. Ethical and aesthetic statements also pose a problem to truth-make theory. For example, the statements "The Mona Lisa is beautiful" or "Murder is wrong" do not seem to have an adequate truth-maker, but seem to be taken as brute facts. Why could not counterfactuals also be taken as brute facts in the way that aesthetic or ethical judgments are? There are many other

23. Molonists are the primary proponents for God containing middle knowledge.
24. William Hasker, *God, Time, and Knowledge* (Ithaca, New York: Cornell University Press, 1989), 30.
25. 1 Corinthians 2:8, Acts 4:27-28, Galatians 1:4 are a few examples.
26. William L. Craig, "Middle Knowledge, Truth-Makers, and the 'Grounding Objection,'" *Faith and Philosophy* 18, no. 3 (2001): 338-339.

example[27] that also illuminate the insufficiency of truth-maker theory to be the ultimate standard of what we consider true or false statements. Proponents of truth-maker theory have acknowledged that some statements may not need truth-makers, but if that is the case, why cannot counterfactual statements be considered part of this excluded group?

As William Lane Craig states, "I think it is evident that anti-Molinists have not even begun to do the necessary homework in order for their grounding objection to fly. They have yet to articulate their ontology of truth, including the nature of truth-bearers and truth-makers . . . Nor have they applied their theory to counterfactuals of creaturely freedom."[28] It seems to me that the Molonist has no reason to reject middle knowledge at this point, until it can be proven that counterfactuals require a truth maker. I myself agree with Plantinga when he states, "It seems to me much clearer that some counterfactuals of freedom are at least possibly true than that the truth of propositions must, in general be grounded in this way."[29] As stated previously in regards to the principle of alternate possibilities, my response here is simply trying to find a reasonable way to maintain both God's foreknowledge of future human actions and human freedom. There is clearly much work to be done in regards to middle knowledge and truth-maker theory. However, at this point, it seems far from determined that middle knowledge is false.

Conclusion: Resolving Richard's Dilemma

As I stated in my introduction, the purpose of this paper was to refute Swinburne's central argument that God's foreknowledge of human actions is incompatible with human freedom. Swinburne claims that God's knowledge must be understood in the "weak sense" which denies God's knowledge of future human actions. In response, I first laid out three reasons why Swinburne's argument is of great importance to the Christian. These three reasons were based on the Christian themes of scripture, sovereignty and salvation. I then proceeded to lay out Swinburne's argument.

Swinburne argued that since humans have freedom of the will, it is possible that a free human action could prove God's foreknowledge to be incorrect. But, since it is a logical contradiction to claim that God's knowledge can be proven wrong, one must deny God's foreknowledge of our actions. However, I took a different route and argued that humans cannot prove God's foreknowledge wrong, while maintaining their freedom.

My argument started by proposing that God in his timeless state possessed knowledge of our future actions based on the truth of what we would do in

27. See Craig, "Middle Knowledge, Truth-Makers, and The Grounding Objection," 341.
28. Craig, "Middle Knowledge, Truth-Makers, and The Grounding Objection," 348.
29. Moreland and Craig, *Philosophical Foundations*, 529.

freedom-permitting circumstances. After the act of creation, God carried this knowledge over into his temporal state. Despite God knowing what we will do in the future, we still possess freedom of the will for two primary reasons. First, his foreknowledge is based on how we would act; therefore, he is not determining our actions. Secondly, we do not require the ability to do otherwise in order to have free actions.

Shortly after my main argument, I responded to two possible objections. These objections were concerned with moral responsibility and with God's middle knowledge. In regards to moral responsibility, I argued in favour of Michael McKenna's principle titled, the L-Reply. In response to the grounding objection, I argued that it is based on truth-maker theory, which seems to be unable to fully account for why God could not possess knowledge of our future actions.

In conclusion, it seems to me that I presented a thorough and well-reasoned rebuttal of Swinburne's argument that God's foreknowledge of future human actions and human freedom are incompatible. Therefore, the Christian has grounds to maintain the strong version of God's omniscience, which includes knowledge of future human actions

Bibliography

Anselm of Canterbury. *The Major Work.* Edited by Brian Davies and G. R. Evans. Oxford: Oxford University Press, 2008.

Boethius. *The Theological Tractates: The Consolation of Philosophy.* Edited by Hugh F. Stewart and Edward K. Rand. London: Heinemann, 1918.

Craig, William L., Paul K. Helseth, Gregory A. Boyd, and Ron Highfield. *Four Views on Divine Providence.* Edited by Stanley N. Gundtry and Dennis W. Jowers. Grand Rapids, MI: Zondervan, 2011.

Craig, William L., and James P. Moreland, *Philosophical Foundations for a Christian Worldview* (2nd ed). Downers Grove, Il: InterVarsity Press, 2017.

Craig, William L. *Time and Eternity: Exploring God's Relationship to Time.* Wheaton, Il: Crossway, 2001.

Craig, William L. "Middle Knowledge, Truth-Makers, and the 'Grounding Objection.'" *Faith and Philosophy* 18, no. 3 (2001): 337-352.

Cowan, Steven. "The Grounding Objection to Middle Knowledge Revisited." *Religious Studies* 39, no. 1 (2003): 93-102.

Frankfurt, Harry. "Alternative Possibilities and Moral Responsibility." *The Journal of Philosophy* 66, no. 23 (1969): 829-839.

De Molina, Luis. *On Divine Foreknowledge: Part IV of the Concordia.* Translated by Alfred J. Freddoso. Ithaca: Cornell University Press, 1988.

Fischer, John M. "Frankfurt-Type Examples and Semi-Compatibilism." In *The Oxford Handbook of Free Will* (2nd. ed.), edited by Robert Kane, 281-307. New York, NY: Oxford University Press, 2011.

Ganssle, Gregory E. "God and Time." In *Internet Encyclopedia of Philosophy.* https://www.iep.utm.edu/god-time/

Hasker, William. *God, Time, and Knowledge.* Ithaca, New York: Cornell University Press, 1989.

Hasker, William. "Explanatory Priority: Transitive and Unequivocal, a Reply to William Craig." *Philosophy and Phenomenological Research* 57, no. 2 (1997): 389-393.

Lewis, Clive S. *Mere Christianity.* New York, NY: Macmillan.

McKenna, Michel, and Derek Pereboom, *Free Will: A Contemporary Introduction.* New York, NY: Routledge, 2016.

Piper, John, Justin Taylor, and Paul K. Helseth, *Open Theism and the Undermining of Biblical Christianity.* Wheaton, Il: CrossWay, 2003.

Swinburne, Richard. "Causation, Time, and God's Omniscience." *Topoi: An International Review of Philosophy* 36, no. 4 (2017): 675-684.

Van Inwagen, Peter. *Thinking About Free Will.* Cambridge, UK: Cambridge University Press, 2017.

Visser, Sandra. "God's Knowledge of an Unreal Future." In *Philosophical Essays Against Open Theism*, edited by Benjamin H. Arbour, 37-44. New York, NY: Routledge, 2019.

Ware, Bruce. *God's Lesser Glory: The Diminished God of Open Theism.* Wheaton, Il: Crossway, 2000.

Saving Deontology Despite Involuntary Beliefs—"Ought" Doesn't Imply "Can"

Trueman Andrews-Gibson
University of Redlands

Abstract

In epistemic situations, we often make statements such as, "You shouldn't have believed that." Such normative judgements seem to rely on an assumption that we choose our beliefs, the view of doxastic voluntarists. According to Kant's doctrine that 'ought implies can,' we cannot apply obligations to anything involuntary. But can we? This paper will provide counterexamples. If people ought to do some things which they cannot, deontological terms such as "permitted" and "obligated" apply to beliefs regardless of whether we control them. Next, this essay explores belief control and refutes a compatibilist defense of free belief, offering important ethical implications.

In common speech we say things like "you shouldn't have believed that," presuming that a person is responsible for her beliefs because she could have chosen otherwise. Philosophers have examined how much control we have over which beliefs we hold. Many have determined that once we think deeply about our belief-formation, we do not seem to make conscious decisions to believe one thing over another. Often, the evidence just guides us to find x more likely than $\sim x$. The relationship between voluntary belief and deontological conceptions is contentious among modern epistemologists. Deontological terms (used interchangeably with "deontic" or "deontology" henceforth), which describe when we are obligated, permitted, or prohibited from believing that x, are useful tools in considering the ethics of belief.[1] It has serious ethical implications to say that sometimes we are permitted to believe one thing and prohibited from believing another. Given that our autonomy over belief is uncertain, deontic terms may not be useful. If this is true, we are foolish to describe a person as praiseworthy or blameworthy for believing anything.

1. William P. Alston, "The Deontological Conception of Epistemic Justification," *Philosophical Perspectives* 2 (1988): 257.

Doxastic voluntarism is the view that people freely elect their beliefs. Proponents of deontology have historically been fond of doxastic voluntarism because of the classic principle that "ought implies can." Arguing that without choice people cannot be obligated to hold any beliefs, some involuntarists, who claim that beliefs are not controllable, have tried to invalidate the use of deontological terms. This essay builds on arguments by Richard Feldman and Sharon Ryan, who have attempted to dismantle Immanuel Kant's famous doctrine. If our arguments are persuasive, the conclusion is this: people do not freely choose beliefs but they do have obligations that ground their methods in the epistemic world. Therefore, even if a person did not choose to believe ~x instead of the truth, x, they can be blameworthy if their epistemic or moral practices were flawed.

We can consider a number of objections made to the principle that "ought implies can" given counterexamples where a person is unable but still obligated to do something. Also needed is a situation that demonstrates an epistemic "ought," as philosophers have typically only used moral scenarios. This will be my first significant addition to the field of belief deontology. Next, I will review William Alston's arguments against doxastic voluntarism. On the philosophy of action, my paper will provide an argument to account for situations where no belief is permissible but action is required. I will show why we can suspend epistemic judgement while responding as if we have formed a belief. The final portion of this essay will address the modern defense of doxastic voluntarism (that is, a free choice of belief). Sharon Ryan takes a compatibilist approach to belief that, I will argue, is unpersuasive in its assessment of how people come to believe propositions. My contribution to the field of belief control will be my arguments against the doxastic compatibilist thesis. Even if Ryan were correct about people having choice of belief, I contend that her account of "choice" makes it irrelevant to any discussion of deontology. I will conclude that while nearly all of our beliefs are involuntary, we still have obligations that should ground our epistemic behaviors. Finally, I will discuss implications for the ethics of belief.

What Is a Choice?

Before delving into my thesis, it seems important to note what I mean by "choice" in relation to discussions about belief. This will be helpful throughout and especially significant when this paper turns to Sharon Ryan's account of choosing beliefs within her compatibilism. When one chooses to believe something, one has the control to choose or not choose to believe that x. There must be two states of mind at two different times if beliefs are indeed chosen: the mental attitude with the intention to believe, and a mental recognition that one now holds that belief. Equally, if it is accurate to say that we cannot choose beliefs, it is the case that (1) one intends to believe that x, and (2) one subsequently does not hold that belief; alternatively, it could be the case that one did not intend to believe that x but one does anyway. If beliefs cannot be chosen,

the former statement explains how people fail to believe things they would like to believe, while the latter describes how they form the vast majority of their beliefs. This understanding of choice should be compatible with any account of belief formation.

Epistemic Justification

An account of epistemic justification, in the context of deontology, will also be useful for the coming philosophical arguments. To say that we are permitted to believe something is also to say that we are epistemically justified. How "justified" is defined is a controversial topic. It may seem natural to hold the justified and unjustified as mirror opposites. A justified belief has sufficient evidence and an unjustified belief is anything without it. However, Alston's use of justified is not that rigorous. We should consider a belief to be justified just when one has not "violated any relevant rules, norms, or principles" by believing.[2] This differs from, say, an evidentialist's perspective that we must suspend judgement on a matter until we have conclusive evidence. While this account keeps emphasis on having good evidence for beliefs, when one has been epistemically responsible and still no answer is completely justified, one is not wrong to believe. The connection between epistemic justification and deontology is as follows: "justified" means that one is definitely "permitted" to have a belief, and "obligated" is a higher standard than "justified." Likewise, a belief one is prohibited from holding (typically used as opposite to "obligated") is one that is not only unjustified, but faces significant contrary evidence and little or no supporting evidence. While it is impossible to assign accurate numeric value to them, we can rank the terminology from weakest to strongest according to belief like so: prohibited, unjustified, permitted, justified, obligated.

Alston's deontological conception of justification helps us distinguish between permissible and obligated beliefs, the latter of which demands the highest standard of evidence. The rules, norms, and principles of permissibility mean going about belief formation with practices that generally produce true beliefs. Perfection is not necessary. While this does not guarantee complete accuracy (what account does?), it is the "most reasonable way to try to have true beliefs."[3] Among these good practices are suspending judgement until seeing credible, unconflicted evidence, verifying with multiple sources, and looking for contrary evidence. An obligation to believe is more—it is a situation where the epistemically responsible person has no excuse *not* to believe.

Take, for example, a woman who begins to suspect that her husband is cheating on her. Sarah has held a strong marriage for a decade, but recently her husband has become newly protective of his cell phone, coming home at a later time de-

2. Alston, "The Deontological Conception," 284.

3. Richard Feldman, "The Ethics of Belief," *Philosophy and Phenomenological Research* 60, no. 3 (2000): 683.

spite no change in work schedule and showing a decreased sexual interest in his spouse. Searching for evidence Sarah finds messages between the husband and a co-worker where certain texts are missing from the conversation. These indicators may permit Sarah to believe that her husband is unfaithful. In forming this belief, she has violated no rules of epistemic justification: she has looked for the evidence available to her, stayed open to the possibility that these signs were misleading, and avoided being blinded by passion. Consider a stronger case. All of the aforementioned signs are present *and* Sarah hires a private investigator to find the truth. She hires a reputable agency which provides a video of her husband in a car with a topless woman, clearly showing their faces and sexual activity. Here the facts break the boundaries of permissibility. It is not only justified for Sarah to believe her husband is unfaithful, it is obligated. Sarah would be nothing short of an epistemic fool to look at those pictures and maintain that her husband has not cheated.

"Ought" Doesn't Imply "Can"

Let us assume that except for some weak influence, we do not have autonomy over beliefs. This, the deontologist worries, seems to imply that we are not responsible for what we believe. The devout Christian cannot be chastised for having no evidence for his world view, Sarah can remain in denial about her husband without ethical consequence, and I can believe that this paper deserves publication without blame. There are a few responses to this concern offered by Richard Feldman and Sharon Ryan with their attacks on "ought implies can." Feldman defends the use of deontological conceptions through the discussion of what he calls "role oughts." That is, "oughts" that are placed on people because of the roles that they adopt. For example, when one becomes a teacher he has an obligation to be patient with his students. In similar fashion, a parent ought to take care of her kids. Even if one finds himself unable to meet all the obligations of his role, he does not renounce those obligations.[4]

A line must be drawn between what one "ought" to do and what a person can reasonably be blamed for doing. We often equate the two, but there is a difference once we realize that people do not lose all obligations just because they are incapable. This distinction has practical applications. As Feldman notes, without ought, a person would not be responsible for changing their future behavior to make it more likely that they *could* meet obligations. A negligent parent, not deserving blame in the short term because she has to work long hours to support her child, still ought to be there for her kid. Therefore, we can excuse her absence to a point but she is obligated to take steps so she can meet what is expected of a parent. From the day she begins neglecting her child, she ought to be a better parent—but starting from day one she is not blameworthy. At some point, say six months later, she has earned enough money and had adequate time to make arrangements so that her child is not

4. Feldman, "The Ethics of Belief," 676.

neglected, however that can be achieved. By distinguishing between "ought" and blame/praiseworthy actions or beliefs, we can explain the dilemma of the negligent parent as violating a moral responsibility, while recognizing that in the short-term she is not blameworthy because it was beyond her control.

Sharon Ryan offers a number of arguments to show why "ought implies can," while intuitive, does not stand up to reason. One issue with the doctrine is that under it an agent could unjustly avoid responsibility by self-sabotaging himself into becoming unable to meet his obligations. The example provided is this: a man (call him Fred) promised to pick his friend up from the airport at 6:00 p.m., but in order to avoid this, he gets on a flight of his own at 5:50 p.m. With "ought" implying "can," it appears that Fred does not have an ethical obligation to pick up his friend after 5:50 p.m.[5] Ryan argues that this account falls short for not capturing that he is responsible, despite being incapable. If Fred can look at his watch at 6:00 p.m. while ascending and feel guilty for not picking up his friend, this illustrates how he is still obligated despite being presently unable. To someone who believes that "ought" must imply "can," he is only blameworthy for getting on a different plane. It seems correct that Fred is also obligated to pick up his friend at 6:00 p.m. because he agreed to. Therefore, Ryan provides a more accurate account of deontological responsibility in this case. Whether due to the role one adopts (like parents) or their previous commitments (Fred), Feldman and Ryan provide counterexamples to the doctrine that "ought implies can." We are left with genuine impossibilities that carry obligations.

"Ought" Doesn't "Can" Even Epistemically

Is there any reason to doubt that just because morally, "ought" does not imply "can," it may epistemically? The articles in my purview have been surprisingly silent on connecting the dots between moral "ought/can" and epistemic "ought/can." It seems that it is simply easier to come up with moral examples when using deontological terms. I will now fill that gap in the conversation about deontology. Consider an example which contains both epistemic and moral elements, and (like Ryan's example) involves putting oneself into a position of incapability. Steve the prosecutor is trying a case against a recurring criminal for armed robbery. The defense counsel brings to Steve's attention the claim by an eyewitness that the assailant was a different race than the indicted man. Not thinking the inquiry is worth his time, Steve presses on and successfully convicts an innocent man. After choosing to ignore that possibility, and looking at the evidence (assume it is pretty strong evidence), Steve could not know that the defendant is innocent while giving his closing argument. However, his poor epistemic practice of ignoring a possibly exonerating witness means that he ought to know that it is not beyond reasonable doubt that he is guilty.

5. Sharon Ryan, "Doxastic Compatibilism and the Ethics of Belief," *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 114, no. 1/2 (2003): 52.

This is not to say that Steve ought to know that the defendant is actually innocent. However, since there was an opportunity to hear contrary evidence, he should know that the defendant is not guilty to the standards of the American criminal justice system. These are standards which Steve has sworn an oath to in his duty as a public servant. Therefore, to describe the defendant as "not guilty" is not the same as saying epistemically, we are justified in believing his innocence. Looking at this situation from an ethical perspective, epistemic standards of justification work in conjunction with Steve's moral responsibilities that come with his role. Unlike the defense, which is one-sided in offering the best representation regardless of the facts, prosecution attorneys have additional responsibilities to see that justice is served. Our justice system accepts the principle of innocent until proven guilty, and prosecutors are obligated to be open to potentially contrary evidence.

At first glance it may appear wrong to say that Steve could not know the defendant is "not guilty"[6] at the end of the trial—after all, Steve *could* have known that the defendant was innocent if he had acted differently. Still, "ought" implying "can" only concerns itself with a particular moment of knowledge and obligation. At the end of the trial, there is no way for Steve to know that the man is "not guilty," but because of his epistemic negligence, as well as the requirements of his role, he ought to know it.

While certain examples offered in the papers by Feldman and Ryan were unconvincing to me, the examples I included, as well as my prosecutor example, offer realistic and epistemically troubling scenarios under the view that "ought" implies "can." Whether an agent is blameworthy will, by virtue of whether he took missteps to arrive at a false belief, depend on each individual situation. The example that strikes me as most powerful is the last. If we do not say that Steve the prosecutor ought to know the defendant is not guilty beyond a reasonable doubt, his incorrect conclusion is epistemically permissible, despite a critical mistake. My argument is that his conclusion is not permissible even though there is nothing telling him, at the end of the trial, that the defendant may be innocent.

Almost All Beliefs Are Involuntary

I will now explore arguments against voluntary belief, the contents of which do not depend on the conclusion that "ought" implies "can." First, William Alston makes note of types of actions (before drawing the parallel to belief) according to how immediate they are to our control. Basic actions, such as raising my arm, are performed entirely at will. There seems to be no figurative room between the decision to perform a basic action and the action's completion. This is

6. Again, I do not mean to equate "not guilty" with proven innocence. By "not guilty," I only mean that the collective evidence does not permit Steve to move forward with his prosecution.

clearly not comparable to how people enter states of belief. I cannot, as simply as raising my arm, start to believe that the Earth is flat. Even if I were offered $10,000,000 for it, I could not elect that belief.[7] The same is true for less direct forms of control. I cannot begin a mental process that will, in one hour or one week, result in my believing the Earth is flat.

A response to this example might be that it is problematic because I have too much previous knowledge about the Earth. After all, I have been told it is a globe and seen documentaries about the planets since I was a child. Perhaps people can choose beliefs when they do not already know something about the topic. Take Rebecca, who knows little about biology. Can Rebecca simply choose to believe that humans share 90% of our DNA with chimpanzees? It is clear that she cannot do this because she's encountered no evidence to support that belief. There is one form of belief control that philosophers commonly accept, which is long-range influence. Rebecca cannot choose to believe how much DNA humans share with a chimpanzee, but she can look into it. If she starts to research genetic similarities between the two species, she is likely to bring herself to form a belief about it. After a few hours of research, Rebecca is justified about her belief on that topic. Observe that she did not form a belief in the way a doxastic voluntarist would claim. She did not make herself believe her original hypothesis that the number is 90% because (unless there was a strange coincidence) the research will show that number to be different. In other words, Rebecca has influenced a belief on the subject of DNA similarity between humans and chimps, but she has not elected to believe anything. The specification of her belief was up to the research she found.

Alston addresses the claim that perhaps we can control beliefs when there is a decent amount of evidence both affirming and denying that x. For a controversial belief, such as the existence of God, there are important considerations in religion, science, and ethics that should guide an internal debate. Assume that religious and nonreligious people make good points while trying to convince someone to agree with them. Does one have a choice here? Alston argues that these situations still do not bring choice of belief because despite the lack of conclusive evidence, "the belief follows automatically, without intervention by the will."[8] Like Simmias to his raft, one clutches the belief that simply appears more likely to be true.

While an evidentialist might claim that we are obligated to suspend judgement on a proposition until there is conclusive evidence, there are many practical situations that call for action when we are not (by the standards of epistemic justification) validated in forming a belief. Alston claims that we form whatever belief seems more likely to us.[9] Under the account of epistemic justification

7. Alston, "The Deontological Conception," 263.
8. Alston, "The Deontological Conception," 266.
9. Alston, "The Deontological Conception," 265.

that Alston and I accept, one can believe that x when his epistemic practices are sound and there is credible evidence in favor of the proposition. When there is evidence for and against x, this account is not satisfied because it violates the epistemic principle of suspending judgement when evidence conflicts. If one looks for contrary evidence and finds it, one becomes obligated to seek additional verification before believing.

Beyond Belief—Hypothesis and Action

What appears to be missing from Alston's perspective is an important distinction to account for actions that are not motivated by epistemically-justified beliefs. In many situations, despite how we act, we should not regard our attitude toward proposition x as a belief, but rather a hypothesis. It is a mistake to assume that because there are situations which require us to act from uncertainty, we must form a belief one way or another. Imagine the situation of thinking I might have forgotten to lock my car as I arrive at class. I usually, by habit, hit the lock button on my key after closing my door, but I do not recall hearing the doors lock shut after I did this. My mind was also pretty occupied on the readings I had just finished, so it is possible I just tuned out the sound. My possible choices, according to a doxastic voluntarist, are to believe that I did lock the door and do nothing or believe that I forgot and run out to lock it. Alston would argue that these are not choices—one will simply appear more likely to me and I will believe one proposition or the other. In fact, I do not have to form any belief about the car. In this situation, I have to act as if I believe something, but that action can just be based on a guess. Of course, an agent in my example does have beliefs such as "my car might be unlocked." When I say that the person does not have a belief, I am referring to the ultimate question being considered, which is whether the car *is* or *is not* locked.

My account of (non)belief and action will use the words "hypothesis" and "guess" for lack of better terms, but I mean to use them in cases where something's truth likelihood is greater than a shot in the dark, say 50/50. In my car's uncertain state of being locked, I can make the hypothesis that it is not locked without actually believing this. Perhaps I run outside and lock the door again, learning that I had actually already done so. When I get back, my friend Lucas asks where I went and I explain. "You thought you didn't lock it?" he asks. "No," I reply. "I had to do something, so I operated as if I had not locked it, but I would not say I believed it was unlocked either."[10] From the perspective of a general fighting a difficult war, it is entirely possible that the enemy will show up here or there, but the general must act. In spite of that action, the general also suspends epistemic judgement when he orders troops to a certain

10. As my philosophy advisor Lawry Finsen pointed out, this example does not necessarily require a guess that the car is unlocked, since one could simply doublecheck without forming any hypothesis on the state of the car. However, we agree that this does not undermine my argument because it demonstrates action in absence of a belief.

location. He is going with his best idea of what is likely. It is not necessarily true that the general believes the enemy will be at that location, but if he does, it is an involuntary and unjustified belief.

At play in these examples is a difference between ordinary and philosophical meanings of "believe." I am arguing for a higher standard, which is when one really judges x to be true. We may use "belief" as a weaker term in conversation because we use "know" when we are more confident. Yet, philosophers usually demand a higher standard of proof than others to say one knows something, so my distinction is not misplaced. This account of belief-action strengthens our understanding of epistemic justification. Indeed, it is wrong to say that a person is justified to adopt a position when both x and $\sim x$ are backed up by evidence not only because the standards of justification have not been met; explicating belief to say one believes that x only when one sincerely finds it true, we see that a person does not choose to form a belief in scenarios of conflicting evidence, even when such scenarios require action. Alston notes that even if it were true that in these cases one chooses to believe x or $\sim x$, that would not provide a foundation for a comprehensive voluntarist account. A limited number of propositions genuinely appear ambiguous, especially when stacked against the natural, unquestioned beliefs people operate on daily.[11] In conjunction with previous arguments against one's ability to choose beliefs, the case for doxastic voluntarism is crippled.

Voluntarism's Last Stand

Not missing a chance to make me qualify with an exception, Richard Feldman demonstrates a type of belief that people seem to have high control over. While I do not see these cases as incredibly relevant to deontology because they affect a petite subset of beliefs, it should be mentioned that some beliefs are controllable. Beliefs about one's immediate surroundings can be manipulated. For example, I have a belief about the lamp at my desk. It is off. I can exercise control over my belief about the lamp by taking action to change it. While I cannot, solely by will, choose to believe that my lamp is on, I can control my belief by thinking, "I will believe that my lamp is on," and then turning it on.[12] Feldman says that controlling these types of beliefs has "prudential and moral significance." Ironically, he gives a rather unrealistic (though not invalid) example to justify that statement. He argues that if the department chair offers a raise to everyone who believes that the lights are on, Feldman better make sure he holds that belief. While this type of belief does seem controllable, it does not refute the involuntarism thesis except in the extreme claim that no beliefs are chosen. What is more important: that you can control your belief about whether your lamp is on, or that you cannot control your beliefs about God, who is a good friend to you, whether your spouse is faithful, or whether your life is meaningful?

11. Alston, "The Deontological Conception," 265.
12. Feldman, "The Ethics of Belief," 671.

It is clear that of all the facets of life that belief touches, one's ability to control beliefs about her surroundings is trivial.

To be fair, Feldman's argument does satisfy my initial explication of what it means to "choose," as related to belief. It seems true that in the lamp example a person (1) intends to believe that the lamp is on, and (2) subsequently believes it is on. For this reason, it appears that I should agree with the proposition that some beliefs are chosen. However, there are reasonable, imaginable exceptions to these situations when an agent will fail to choose his or her belief. This is because (2), that the agent believes the lamp is on, can go wrong. You could find that after having an intention to believe that the lamp is on, you flip the switch and find the bulb dead. Manipulating one's immediate surroundings is, it seems, easy enough that exceptions like these should not disqualify Feldman's argument from constituting a choice to believe.[13] Likewise, the fallibility of these choices, as well as the small number of propositions they apply to, confirm my previous conclusion that involuntarism is correct. The vast majority of beliefs are not chosen.

Does Involuntary Stand to Ryan's Objection

I will now refute a modern theory of doxastic voluntarism. Sharon Ryan argues for doxastic compatibilism, which seeks to combine choice of belief with an "appreciation of our evidence." If her view is accurate, involuntarists' emphasis on the evidence swaying our beliefs is compatible with voluntarists' emphasis on choice.[14] Ryan admits that we can hardly ever "control the fact that our beliefs will be determined by our awareness of the evidence in much the same way that we cannot control the fact that our actions will be determined by the laws of nature." For that reason, this defense has already revealed itself to be quite different from the original claims of doxastic voluntarism. More importantly, it will not help us understand deontology better than involuntarism, which argues against belief control. What philosophers have generally meant by pondering the question of whether we choose beliefs is whether we can control them by will. Ryan's coming description of "choice" makes her account extraneous to the deontic question of whether terms like "permitted" and "obligated" should be applied to beliefs. If people "choose" but do not control their beliefs, they are still subject to belief-destiny in the way that determinists hold that one's will cannot dictate one's life. Therefore, it is more accurate to describe beliefs as not being chosen.

Ryan uses the word "intentional" to mean doing something purposefully, but

13. I suspect there may be additional cases where, regarding the belief that forms in reaction to perceived reality, one's ability to manipulate beliefs is similar to Feldman's argument. This means there is the potential for other beliefs to be controllable. However, determining where one draws the line between reasonable and fallible "controlled" beliefs (since manipulating reality, and subsequent beliefs, can always go wrong), will be difficult.

14. Ryan, "Doxatic Compatibilism," 70.

not necessarily consciously. People form beliefs intentionally, she claims, but not with the "explicit intention" to take on any doxastic attitude. She gives the example that she is typing and intentionally hitting each letter to form her sentences, though she is typing "unconsciously, automatically, and purposefully."[15] Ryan takes examples of action she regards as free before explaining why doxastic attitudes are similar to these actions. For example, one's laughing at a funny joke is caused by an "appreciation of the joke [and an] intention to laugh." One cannot avoid laughing when she finds something funny. One also cannot stop believing that a car is hurling at her if that is what she is witnessing. That does not mean that she has no "control over what [she] does or what [she] believe[s]." Ryan can laugh at a good joke and believe that a car is about to hit her freely. Her view is that it is free because she is believing what she means to believe. It did not "just happen" to her.[16]

Unfortunately, it appears that these scenarios do not actually demonstrate freedom in any meaningful sense. Just because one intended to laugh at a good joke does not mean that she chose to do so, since she could not have chosen otherwise. While "ought" may not imply "can," "chose" does imply that a person could have "not chosen." Likewise, while Ryan claims that her belief concerning a car about to hit her was what she meant to believe, that is irrelevant to the fact that she could not have chosen to believe otherwise. It is also not necessarily what a person in that situation would mean to believe. As the proverb goes, if our worst fears lie in anticipation, a person might not wish to believe that she is going to be hit by a car, particularly if the accident will be fatal. It seems equally incorrect to describe the attitude of a person, about to be hit by a car, as "choosing." The belief of impending doom seems to "just happen" to the person.

So What, Ryan?

Not only is the compatibilist approach to belief wrong in my view, but I maintain that even if it is correct, it does not aid deontology's search for a meaningful approach to the ethics of belief. If we are still slaves to our perceptions and the evidence available to us, our "choice" to believe things intentionally is not relevant to my inquiry about control. This leads us to accept the reality of involuntarism—we don't elect the vast majority of our beliefs. Feldman's contribution disproves the extreme view that no belief is chosen, but it is insignificant to deontology's application to belief. We also have the fact (one that I do not understand any contemporary academic to contest) that we have influence over many beliefs, particularly regarding topics that we do not know much

15. Ryan, "Doxastic Compatibilism," 71. This section of Ryan's essay seems to rely too much on synonymity for definition.

16. Ryan, "Doxastic Compatibilism," 73-74. I quote a lot here, as I do nowhere else in the paper, for the purpose of getting Ryan's argument in her own words. I find her distinctions a bit odd at times, and want to prevent misrepresenting her account in my critique.

about. I brought myself to form many new beliefs about philosophical issues when I started taking classes in it. Of course, I did not exercise control over what exactly those views entail.

This does not prevent a meaningful approach to the ethics of belief because deontological terms like "permitted," "prohibited," and "obligated" apply irrespective of our control. Alston's argument that deontology is undermined because beliefs are not chosen has been disproven by Feldman's "role oughts" and Ryan's argument against "ought implies can," as well as my extension to epistemic "oughts." There are roles, as well as situations and epistemic behaviors, that place obligations on people. There are practical reasons to describe one as maintaining an obligation even when they are currently unable to fulfill it. This is the impact that obligation has on future epistemic and moral behavior. A parent, no matter how overworked, has an obligation to take better care of her neglected child. It is also accurate to say that Fred, who agreed to pick up his friend from the airport, is not excused because he got on a plane himself. He is still required to pick up his friend even after getting on a different plane. Steve the negligent prosecutor ought to know that his target is not guilty to the standards of the American justice system even though he lacks evidence to guide that conclusion. Denying that obligation would not require Steve to change his poor epistemic practices for the future, as he should be open to contrary evidence.

An Approach to Ethical Belief

People should not be excused from poorly informed beliefs just because they did not choose to have them. To do so would have disastrous ethical implications—people could go through life with great ignorance, absent of responsibility. Additionally, people could take steps to avoid being informed just to dodge epistemic obligations that come with knowledge. Because we are permitted to have some beliefs and prohibited from having others, we should continue to make statements akin to "you shouldn't have thought that" in casual conversation and academia. When someone has a foolish belief that naturally appears true to him, and he has not sought justification or questioned his premises, we can disapprove of that person's epistemic practices as well as his incorrect belief. Individuals who have any interest in believing what is true and disbelieving what is false must consider their epistemic assumptions, actions, and requirements for justification. Deontology is important in a world where people are often too presumptive and reckless in their epistemic practices, allowing truth-seekers to maximize the accuracy of their beliefs.

Bibliography

Alston, William P. "The Deontological Conception of Epistemic Justification." *Philosophical Perspectives* 2 (1988): 257-299.

Feldman, Richard. "The Ethics of Belief." *Philosophy and Phenomenological Research* 60, no. 3 (2000): 667-695.

Ryan, Sharon. "Doxastic Compatibilism and the Ethics of Belief." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 114, no. 1/2 (2003): 47-79.

Aristotle on Self-Indulgence and Incontinence

Angelo Chan Borges
Cornell University

Abstract

In this paper, I try to give a plausible interpretation of Aristotle's discussion on self-indulgence and incontinence in Book VII of *Nicomachean Ethics* (NE). The ambiguity in Aristotle's conception of vice is reflected in the long tradition of scholarly dispute on whether Aristotle's vicious agent is a principled follower of a wrong conception of the good. Answering in the negative, Jozef Müller has recently provided a promising reading of the vicious agent as a follower of mere appearances who lacks any stable conception of the good. However, arguing against Müller's reading, I shall show that Aristotle's discussion of self-indulgence and incontinence is intelligible only if the vicious person in Book VII is principled. Then, having argued in favor of the conception of the principled vicious agent, I go on to explain how we can make sense of Aristotle's view that the incontinent person is better than the self-indulgent person in terms of their susceptibility to persuasion.

Whether self-indulgence or incontinence is a more desirable state is an interesting inquiry that Aristotle makes. In Book VII of *Nicomachean Ethics* (NE), Aristotle frames his discussion on this topic as a puzzle that offers a position to which he then responds. Aristotle's discussion is by no means clear-cut. Yet Jozef Müller's recent challenge to the standard view of vice, on which Aristotle's vicious agent is principled rather than conflicted, seems to have only made the task of understanding Aristotle's puzzle and his answers to it more difficult. Thus, in Section I of this paper, my aim is to defend the standard view of vice against Müller's proposal of the conflicted vicious agent, arguing that it is only on the standard view that we can properly make sense of Aristotle's puzzle concerning self-indulgence and incontinence. Then, in Section II of this paper, I take on board the standard conception of vice, and in doing so, I try to argue for my own interpretation of Aristotle's solution to the puzzle, showing exactly how, on Aristotle's view, self-indulgence is worse than incontinence because only the latter state is curable by means of persuasion or argument.

I. Defence of the Standard View

To begin, let us compare the two different scholarly readings of vice in NE.[1] On the standard view, the vicious person has an incorrect but stable conception of the human good according to which she organizes her life, "just as the virtuous person does with the correct one."[2] In this way, since Aristotle says self-indulgence is concerned with bodily pleasures (and pains) [1148a18-20], the self-indulgent person is "one who has adopted a general policy to pursue bodily pleasures," that is, "a principled pursuer of pleasure."[3] Arguing against this view, however, Müller suggests that the vicious person is conflicted, because she follows mere appearances of pleasure (which are often deceiving) so that her actions are based on current interests and feelings rather than any stable, rationally grounded principle (or reasoned conception of the good).[4] As such, she is described as "the fool who although capable of reasoning and decisions does not organize her life around some aim or purpose . . . but wanders from one random pleasure to another, indulging in any that is at hand."[5] There are, then, two very different views of vice.

To illustrate an important feature about Müller's view of vice, I look at how it accounts for the vicious person's regret. After all, Müller's view is attractive because it can reconcile Aristotle saying the vicious person has "no regrets" in Book VII (1150b29) and also that she is "full of regret" in Book IX (1166b25). Müller argues that when the vicious agent is acting in pursuit of what now appears most pleasant to her, she has no regrets at the moment of action. But when these actions lead to subsequent pain for her at a later time, and because she has no stable conception of the good, she comes to regret doing what she previously enjoyed.[6] The important feature here is that the source of the vicious agent's regret, namely, her subsequent pain from doing the action, does *not* involve her identification of the normative reasons that make the action bad.[7] Thus, for example, this agent might have severely beaten his mother for money, and when she eventually falls into a coma, he would not be one to regret making his mother comatose for the reason that this is simply a bad thing to do, for he is merely pained that he can somehow no longer extract money from her. Müller argues for this reading of regret because he says it is in this specific sense that the vicious agent has no regrets in Book VII: "She does not regret the harm to others or disapproves—from the point of view of the normative standards of virtue—of the way she is."[8] But the vicious agent does have regret, as in Book IX, in the sense that "she regrets that things did not work out for her as she

1. Until noted otherwise, all translations are from Aristotle, *Aristotle: Selections*, trans. Terence Irwin and Gail Fine (Hackett Publishing, 1995).

2. Josef Müller, "Aristotle on Vice," *British Journal for the History of Philosophy* 23, no. 3 (2015): 460.

3. Müller, "Aristotle on Vice," 460.

4. Müller, "Aristotle on Vice," 464.

5. Müller, "Aristotle on Vice," 465.

6. Müller, "Aristotle on Vice," 467.

7. Müller, "Aristotle on Vice," 467.

8. Müller, "Aristotle on Vice," 468.

had hoped for or that something else were not pleasant to her."[9]

Though this feature of Müller's view has been important in trying to resolve an apparent contradiction in Aristotle's accounts, I shall show that it is also a basis for rejecting his view. To do this, I now take up Aristotle's puzzle on self-indulgence and incontinence. First, Aristotle explains, the self-indulgent person is one who "acts to pursue what is pleasant because this is what he is persuaded and decides to do," whereas the incontinent person "acts not because of rational calculation but because of [his feelings]" (1146a31-34). Now, the puzzle is this: the claim that self-indulgence is a better state to be in than incontinence because the former state is more curable (1146a31-35). Aristotle does not explicitly say why this claim is puzzling—he seems to assume that our intuition or experience would suggest otherwise.[10]

Nonetheless, Aristotle goes on to state the reasoning behind the claim, setting it up to be refuted later. Thus, according to the puzzle, the self-indulgent person "is the easier to cure, because he *might be* persuaded otherwise" (1146a33-1146b3; emphasis added). By contrast, persuasion is ineffective on the incontinent person, because "though already persuaded to act otherwise, he still acts wrongly" (1146a33-1146b3). At this point, I emphasize two things about this puzzle. First, Aristotle considers persuading the self-indulgent/incontinent to act otherwise in such a way that cures her (as opposed to merely persuading her to act otherwise). Thus, while the persuasion here concerns particular actions, its aim is to cure the agent.[11] Second, "curing" an agent should result in some change of disposition or state that would be more stable (as a state) than changing appearances of the moment.[12]

Having described Aristotle's puzzle, I argue that it cannot be adequately formulated under Müller's view of vice. Suppose, then, the self-indulgent agent in the puzzle to be Müller's fool, who acts on mere appearances rather than any reasoned conception of the good. It is not difficult to see, on first impression, that persuasion is no more effective on the fool than it is on the incontinent person. For insofar as Aristotle is concerned with the sort of persuasion that cures the agent, the goal of curing implies that the persuasion must not only get the agent to do the right action, but also get her to do it*for the normative reasons that the action is good*. But from what we have said, it is clear that the fool cannot be persuaded in this way. For, on the one hand, the fool follows

9. Müller, "Aristotle on Vice," 468.

10. For instance, the claim would entail that a person who enjoys unfaithful sex and believes it to be good, is better than the person who thinks unfaithful sex is bad but does it anyway. This might seem puzzling.

11. I take "curing" to mean at least some improvement in the disposition of the agent, not just complete recovery. After all, Aristotle thinks a virtuous person must enjoy doing the right actions she does, but his account of habituation implies that persuasion alone cannot get someone to take pleasure in doing some correct action, since the person must actually do the action before he can come to enjoy it (1103b7-1103b26).

12. I owe this point about curing implying some sort of stability to Rachana Kamtekar.

mere appearances and so will not be receptive to arguments concerning some reasoned conception of the good. On the other hand, arguments based on mere appearances can get her to do the right thing but not also for the right reasons. So the fool's vicious disposition of following mere appearances of pleasure will not be improved in any way, even if she may "appear" to be better at the moment of acting rightly. Consider, for instance, the mother-beater who will be persuaded not to hit his mother *just so* he can extract more money from her.

Müller himself admits as much when he explains that his vicious agent "is not a competent listener of ethical philosophy," not because "the arguments about why certain actions are virtuous will fall on deaf ears" due to her "fundamental disagreements about the very facts of ethical life," but because "she is not even at the stage at which such disagreements (or agreements) can arise."[13] Consequently, since one can readily see that Müller's fool is not any more persuadable, and so not easier to cure in this way, than the incontinent person, the claim—that self-indulgence is better than incontinence—becomes so implausible that Aristotle, with seemingly nothing left to solve, hardly has a puzzle at hand.

The obvious thing to do now is to show that the standard view of vice does not similarly succumb to the pitfall of Müller's view of vice. Recall the puzzle's claim that it is possible to persuade the self-indulgent person to act rightly in such a way that cures her (1146b34-35). If the self-indulgent person acts out of a stable, albeit incorrect, conception of the good, the claim seems plausible, because we may then presume that her incorrect conception of the good only has to be corrected in order for her to start acting rightly for the right reasons. After all, it would seem like she is a follower of reason who has been led astray by having the wrong beliefs. Of course, this presumption might be wrong after we are given a further picture of her moral psychology, perhaps relating to her feelings and pleasures. But the point here is that the observation alone that she is the sort of person to act out of a stable but incorrect principle, absent anything more, is at least not clearly inconsistent with the possibility of persuading her to do otherwise. By contrast, for Müller's fool, the observation alone that she is a follower of mere appearances was sufficient for seeing the ineffectiveness of rational persuasion. My point is well taken if we reflect on the fact that we do not expect the ordinary person to have a flawlessly correct conception of the good, and yet we generally expect her to be able to change her actions upon being persuaded by another who offers her the correct reasons for acting. We can perhaps ask questions like "how wrong does one's conception of the good have to be in order for one to be 'too far gone' to be persuaded otherwise?", but this just shows that there are uncertainties and questions left to be resolved. As I see it, then, it is only under the standard view of vice that we can adequately make sense of Aristotle's puzzle, and this being so, the vicious person in Book

13. Müller, "Aristotle," 476; emphasis added.

VII must be principled.

Finally, I cite one last glaring piece of evidence in support of the standard view of vice. When Aristotle compares the incontinent person with the vicious person later on, he uses a city as an analogy: "[T]he incontinent man is like a city which passes all the right decrees and has good laws, but makes no use of them . . . [whereas] the wicked man is like a city that uses its laws, but has wicked laws to use" (1152a20-24).[14] Plainly, this wicked city not only fits better with but indeed can only fit with a principled vicious person. For just as the wicked city possesses and uses laws (and the laws suggesting some sort of stability), so too the vicious person possesses and acts on some stable, rationally grounded principle. Further, since it is obvious that the virtuous person would correspond to a city possessing and using good laws, the analogy implies that the vicious person would be the exact opposite of the virtuous person, which can only be true if the vicious person is principled. This mirror image between vice and virtue, then, rules out the potential reply on Müller's part that his vicious agent, despite not having a stable disposition to act in a certain way, has a "settled disposition" that can correspond to the wicked city's laws.[15] Alternatively, we might consider that if Aristotle really intended something like Müller's conflicted vicious person, the city this person resembles would instead be a disordered one in which there are *no* laws and yet wicked actions run rampant, and so Müller ought to explain why Aristotle have not done this. Both the puzzle and the analogy, then, show that there are glaring difficulties on Müller's view that, if his view is adopted, would render us unable to account for these smaller but nonetheless constituent parts of Aristotle's account in Book VII.

II. Aristotle's Solution to the Puzzle

Having argued that the standard view of vice is the only proper way to understand the puzzle, I now take on board this conception of vice and focus on Aristotle's answers to the puzzle. Aristotle starts by more clearly defining the incontinent (without qualification) and the self-indulgent: while both of them pursue bodily pleasures and avoid bodily pains to excess, the latter does so "by choice" and the former "contrary to his choice and his judgement" (1148a18-22). He goes on to argue against the ranking proposed in the puzzle on the basis of two different considerations: C1) who is capable of acting more viciously (both frequency and magnitude) and C2) who is more curable. In C1, Aristotle points out that the self-indulgent person is worse because she does vicious acts even when she has weak to no appetite (1150a20-b6), suggesting that she has less restraint and so will act badly more frequently than the incontinent person.

14. For the remainder of the paper, the translations are from Aristotle, *Complete Works of Aristotle, Volume 2: The Revised Oxford Translation*, ed. Jonathan Barnes (Princeton University Press, 1984).

15. For Müller's discussion of his vicious agent's settled disposition, see Müller, "Aristotle," 470.

Furthermore, he also asks us to consider what the self-indulgent person would have done if she had strong appetites (1150a20-b6), suggesting that the nature of her acts would be more vicious. Then, appealing to the same consideration of curability made in the puzzle, in C2, Aristotle maintains the position that the self-indulgent person is worse, because, in fact, the self-indulgent person is incurable, while the incontinent person curable.

But whereas what Aristotle meant in C1 seemed pretty clear, we shall see that there are some difficulties with understanding C2. The reason that Aristotle thinks the self-indulgent person is incurable is that she pursues the wrong things for their own sake by choice, and so has no regrets, and a person "without regrets cannot be cured" (1150a16-b6). However, it is not entirely clear what Aristotle means that the self-indulgent agent has no regrets and that she is incurable. On the question of regret, since we have agreed that the self-indulgent person is one who has adopted a general policy to pursue bodily pleasures, we may understand Aristotle as saying that, in addition to not having regrets at the moment of action, the agent will also not come to regret her actions. Then, on the question of curability, one might wonder whether the self-indulgent agent is incurable without qualification or insofar as she is unreceptive to rational persuasion? I suggest that the latter interpretation is correct because Aristotle frames his answer of the incurability of the self-indulgent agent in the terms, "[t]his is why the position is not as it was expressed in the formulation of the problem" (1150b29-35), and we have already seen that the puzzle was solely concerned with persuasion as a means to curing. And indeed, consistent with this interpretation, Aristotle later goes on to explain that, since the self-indulgent is convinced to pursue excessive bodily pleasures but the incontinent is not, "it is on the contrary the [incontinent] that is easily persuaded to change his mind, while the [self-indulgent] is not" (1151a11-15).

While it may be granted that curability is limited to persuasion, it still remains to explain how on Aristotle's view, contrary to the puzzle, persuasion does not work on the self-indulgent but might work on the incontinent. To answer this, let us examine Aristotle's description of the self-indulgent agent's conviction, on which he acts to pursue excessive bodily pleasures. First, Aristotle says that the self-indulgent agent's conviction "destroys the first principle" (1151a11-1151a19). Since, on the standard view of vice, first principle means the good principle,[16] Aristotle seems to be affirming here that the self-indulgent's conviction makes him unable to entertain right opinion about the conception of the good, implying that he will not be changed by argument. Second, Aristotle describes the origin of the self-indulgent agent's conviction, that is, how he came to be convinced to pursue bodily pleasures, by appealing to the agent's character.

At this point, then, let us briefly examine Aristotle's account of character. In

16. See Müller, "Aristotle," 464, footnote 10, for this suggestion from Müller.

NE II, Aristotle explained that characters pertain to the part of our soul that lacks reason but that is receptive of and able to partake in reason (1102b10-1103a5), and that a virtue of character is a stable disposition to act and feel in a certain way, acquired through habituation, that is, performing actions of a certain type (1103b7-1103b26). Thus, to acquire a temperate character, for instance, one would have to engage in temperate actions, experiencing what is good about these actions and coming to take pleasure in them, so that one can choose these actions for their own sake. Later, in Book VI, Aristotle adds that virtue of character makes the goal correct, while intelligence (knowledge that guides one's actions towards the human good) makes the actions done in pursuit of the goal correct (1144a8-10).

Now, then, we come back to our discussion of the self-indulgent agent's incurability by persuasion. Aristotle says, "the self-indulgent man is convinced *because he is the sort of man to pursue them*" (1151a12-13; emphasis added.) and "neither in [virtue nor vice] is it reason that teaches the first principles" (1151a17-18). In other words, then, the self-indulgent agent is convinced to act in pursuit of bodily pleasures because of his self-indulgent character, and it is this character, as opposed to his reason, that teaches his incorrect principles of action. Also relevant here is Aristotle's claim that full intelligence requires virtue of character (1144a31-32), which suggests that there are some judgments guiding actions that one can only come to make by performing some relevant type of actions and experiencing what is good about them. Since, as mentioned earlier, Aristotle does say that the self-indulgent person excessively pursues bodily pleasures "by choice, for their own sake and not at all for the sake of any result distinct from them" (1150a16-a30), we have further confirmation that she has a vicious character. Thus, piecing all these ideas together, Aristotle seems to want to say that the self-indulgent agent is resistant to persuasion because, by experiencing what is good about and taking pleasure in self-indulgent actions, he came to form value judgments, unique to his character type, that are fundamentally skewed or corrupted, so that upon hearing some argument contrary to these judgments, he shall indeed listen only with a deaf ear. For, since "vice is unconscious of itself" (1151a1-5), the self-indulgent agent is oblivious to the fact that he is doing wrong. Contrary to the presumption made in the puzzle, then, the state of self-indulgence is not merely a matter of holding the wrong beliefs, because the self-indulgent agent's conviction to do the wrong thing, which destroys the good principle and with which he resists persuasion, is rooted in that disposition concerned with pleasures and feelings, namely, his character.

The final thing to account for, then, is how persuasion, insofar as it is for curing, might be effective on the incontinent person. The only elaboration we get from Aristotle on this point is an analogy: "wickedness is like a disease such as dropsy or consumption, while incontinence is like epilepsy; the former is a permanent, the latter an intermittent badness" (1150b29-1151a1). Aristotle's solution, then, seems to rely on the fact that the incontinent person's feelings do not always conflict with her correct judgment, since incontinence is an intermit-

tent condition. So, while Aristotle concedes to the puzzle that persuasion shall be ineffective on the incontinent person when she is overcome by her pleasures, he argues that persuasion will work at other times when she *is* able to act on her judgment. But we might worry what exactly the incontinent person has to be persuaded of when she is not overcome by feelings—after all, has she not already decided on the right action?[17] To answer this question, let us recall that the "the best thing in [incontinent person], the first principle, is preserved" (1151a23-24). This means that her principle of actions is already correct, so this is not what she has to be persuaded of. The remaining possibility, then, is that she is persuaded to do particular actions. Thus, given Aristotle's appeal to character, it seems like he wants to say that, insofar as the incontinent person possess the correct principles but fail to identify the relevant sort of actions, it will be possible to persuade her to do actions of the type that will help her to take pleasure in the right things, training up her virtue of character, so that, in time, her pleasures will become more disciplined, so that she can come to successfully do some virtuous action (or refrain from doing some vicious action). This is, then, a matter of lacking particular knowledge about the sort of actions that will improve her character, and it is easy to see how the persuasion that relates to conveying this knowledge will not be effective on the self-indulgent person. But still, if this view is taken, the implication we get then is that the incontinent person who can effectively identify the relevant particular actions to improve her character, will not need any external persuasion and she can, in this sense, effectively cure herself. This implication may seem problematic, especially since Aristotle does not seem to suggest this kind of idea anywhere else in his account. But while I note this difficulty, this view is at least not inconsistent with Aristotle's idea that (external) persuasion is not necessarily, but rather only possibly, effective on the incontinent person.

In this paper, I have focused much of my attention on unravelling the interpretational difficulties surrounding Aristotle's discussion on incontinence as better than self-indulgence. Along the way, however, I also challenged Müller's view of the conflicted vicious person by pointing out how its adoption will threaten the intelligibility of Aristotle's puzzle on the same discussion. In doing this, then, I diverged from the long tradition of the scholarly dispute revolving around Aristotle's view of vice. For, rather than resting my position on grand ideas like the vicious person's regret, my approach has instead been limited to a much more minor aspect of Aristotle's account, namely, a particular puzzle. Nonetheless, given what other scholars have already contributed towards the debate for the standard view of vice, this addition might prove to be quite useful as it shines light on the possibility that there might be a lot more aspects of Aristotle's account in Book VII, on the same scale as the one I have considered, that will become quite unintelligible upon adopting the view of a conflicted vicious person.

17. I thank Rachana Kamtekar for pointing out this important difficulty.

Bibliography

Aristotle. *Complete Works of Aristotle. Volume 2: The Revised Oxford Translation.* Edited by Jonathan Barnes. Princeton University Press, 1984.

———. *Aristotle: Selections.* Translated by Terence Irwin and Gail Fine. Hackett Publishing, 1995.

Müller, Josef. "Aristotle on Vice." *British Journal for the History of Philosophy* 23, no. 3 (2015): 459-477.

Alternatives to Consent

Christina Ge
Brown University

Abstract

Consent is often epitomized as a requirement of sex. Yet, is consent sufficient for fair and just sex? Consent in all other aspects of life—consenting to surgery, consenting to a cavity filling—is by nature, passive. Passive acceptance of sex allows for coercion, harmful ambiguity and self-regulation. There exists a wide range of sex that lies between enthusiastic and rape, and I argue that consent is inadequate as a model for equitable sex. I explore two alternative models that position desire and negotiation at the forefront of "just" sex instead of consent. Ultimately, I propose a combination model of consent, desire, and negotiation that highlights my conception of what just sex should be. There remains much work to be done in this field, and I hope this essay contributes to emphasizing women's subjectivity and desire in sex.

Introduction

> "Looking at him like that, so awkwardly bent, his belly thick and soft and covered with hair, Margot recoiled. But the thought of what it would take to stop what she had set in motion was overwhelming . . . It wasn't that she was scared he would try to force her to do something against her will but that insisting that they stop now, after everything she'd done to push this forward, would make her seem spoiled and capricious, as if she'd ordered something at a restaurant and then, once the food arrived, had changed her mind and sent it back."[1]

The fictional *New Yorker* piece "Cat Person" describes unwanted sex between Margot, a college student, and Robert, an older man she met at the theatre. Many readers experienced a relatable discomfort in this depiction of consensual but highly undesirable sex. The excerpt above realistically captures the point at

1. Kristen Roupenian, "Cat Person," *New Yorker*, December 4, 2017, https://www.newyorker.com/magazine/2017/12/11/cat-person.

which many women believe it is "too late" to refuse a sexual encounter. While consent is widely accepted as the requisite for sex, it does not seem adequate to achieve "just" sex. Margot's case poses a challenging philosophical question for consent, which alone, cannot prevent problematic sex. While consensual sex is not rape, there exists a wide range of sex that lies between enthusiastic and rape. For these in between cases, there is no widely accepted standard to achieve ethical sex, or the mutual respect and enjoyment of partners. In this paper, I will first outline consent's shortfalls in "gray sex" through Nicola Gavey's work. Then, I will evaluate two alternatives to consent: Ann Cahill's framework of desire and Rebecca Kukla's language of sexual negotiation. Lastly, I will conclude that while both alternatives have potential shortfalls, combined with consent, these parts of a whole together form a robust conception of what just sex should be. While just sex is something to aspire towards in all sexual contexts, I will mostly examine the conception in heterosexual sex, which have influences on gender and sexuality that go beyond the scope of this paper.

To explain what "gray sex" is, Gavey describes a series of consensual but problematic sexual encounters that are not rape, sexual violence, or coercion through interviews in "Technologies and Effects of Heterosexual Coercion." Heterosexual women, like Margot, conveyed passivity in their sexual experiences by merely "going along with it" as if separated from the act. One woman said she detested the sex but did not feel it was her right to stop it. Another woman felt like she had led on the man and would be irresponsible if the encounter did not end in penetrative sex. The male partners hold entitled expectations, as one man even said: ". . . if you didn't want to make love, why did you come?"[2] The interviews convey a general sense of not "having the language to say no," and even if so, an intuitive understanding that a woman's refusal would have little effect.[3] In other cases, Gavey observes that women felt as if something was abnormal about them if they did not enjoy the unwanted sex. In other cases, they felt obliged to "give" sex as an act of generosity or thanks.[4] Gavey argues that these sentiments stem form society's disciplinary and normative pressures, causing women to self-regulate their behaviour in favour of men's sexual interests.[5] In one case, Pat even says that her partner would have stopped if she had yelled for him to stop, but she would never have asserted herself that aggressively.[6] While men are not necessarily forcing women to have sex, women still feel coerced due to heterosexual pressures that shape their behaviour. Often, they feel a responsibility to not let the man down, and to set aside their own feelings to achieve the male orgasm. Gavey understands these encounters as a "complex gray area" between mutually consenting sex on one hand, and

2. Nicola Gavey, "Technologies and Effects of Heterosexual Coercion," *Feminism Psychology* 2, no. 3 (October 1992): 334.
3. Gavey, "Technologies," 334.
4. Gavey, "Technologies," 335.
5. Gavey, "Technologies," 337.
6. Gavey, "Technologies," 337.

rape or sexual coercion on the other.[7] To Gavey, there exists "a continuum of sexual violence," and a key distinction should be drawn between consensual and *wanted* sex.[8] The spectrum looks something like: sex, gray sex, and rape. Consent is not sufficient in ensuring just and pleasurable sex, and some other measure must be in play. Before exploring alternative frameworks, I will take a moment to outline the deeper concerns underlying these gray cases.

II. What's Wrong with Gray?

First, Gavey points out that consent inherently incorporates dangerous heterosexual norms that lend themselves "as a cultural scaffolding for rape."[9] Namely, within consent, the "male sexual drive discourse" poses men as needing sex and women as agents to give men sex to maintain certain desired relationships or outcomes; Gavey calls this the "have hold discourse."[10] Insofar as men are the dominant creatures with sexual needs, female sexuality only matters when it aligns with male desire. As such, women are conditioned to understand their sexuality as something that can be offered and exchanged for emotional stability, pacification, or simply peace of mind, with no value on its own. In this sense, the discourse is not about maximizing women's sexual desire or pleasure but merely giving them an option of saying "yes" or "no" to men. In practice, men often "request" sex while women agree or refuse. Kukla's inclusion of Joan McGregor's quote is helpful: "Consent is normatively significant since it is the method by which we grant others a right to cross our intimate borders."[11] In the same way that people "consent" to surgery or the treatment at dentist appointments, women consent to sex with similar passivity and acceptance. As Kukla notes, the consent model requires the "acquiescence of one person (implicitly assumed to be a woman) to someone else's (presumed to be a man's) requests."[12] While consent is not inherently problematic, due to the problematic influence of heterosexual norms, it subjects women to instruments of male pleasure and are subsequently, deprived of sexual agency.

Second, Gavey challenges whether the consent model is enough to distinguish between coerced or un-coerced sex. As Cahill observes, consent (which can shift depending on feelings or desire) is a slippery slope that may be subject to change at any stage during the encounter; the pressures around self-regulated consent can look similar to coercion, at which point the gray sex resembles rape.[13] In effect, as Cahill observes, Gavey's interviewees struggle to convey

7. Ann J. Cahill, "Recognition, Desire, and Unjust Sex," *Hypatia* 29, no. 2 (Spring 2014): 307.
8. Gavey, "Technologies," 347.
9. Gavey, "Technologies," 347.
10. Cahill, "Recognition," 307-308.
11. Rebecca Kukla, "That's What She Said: The Language of Sexual Negotiation." *Ethics* 129, no. 1 (October 2018): 75.
12. Kukla, "That's What She Said," 75.
13. Cahill, "Recognition," 307.

their experiences as sexual violence, and one method to prevent being raped is to refrain from defining it as rape.[14] In these cases, consent is not enough to help women understand their experiences as incidents of sexual violence when much of their regulation is self-inflicted, which prompts women to feel responsibility for their own discomfort. Once consent is no longer understood as a definite "yes" or "no," but rather, as a fluid set of feelings and circumstances that can be subject to change at any moment, consent, as a structure, is insufficient to achieve a positive framing of mutually desired sex. Thus, the complexities in the gray area between rape and sex cannot be captured in terms of consent. As Kukla points out, consent at its best is "completely autonomous [and] unmanipulated," which is still insufficient for sex to go ethically well; she observes that "we can consent to all sorts of lousy sex, including demeaning, boring, alienated, and unpleasantly painful or otherwise harmful sex."[15] Therefore, consent remains necessary for ethical sex but not sufficient.

III. Desire as an Alternative

Cahill answers Gavey's problem of distinguishing between rape and gray sex through desire. To Cahill, sex without desire raises "an ethical red flag."[16] Gavey's situations seem wholly unethical as only one actor is experiencing sexual desire and imposing that upon the undesiring other.[17] Women must desire sex in order for the act to be autonomous, as opposed to Gavey's examples of "giving" sex due to societal pressures. Desire, rather than consent, places women in a position of sexual subjectivity rather than acquiescence.[18] As such, Cahill argues that desire of both parties should be required, replacing consent as a necessary condition for ethical sex. Only when women's sexual subjectivity is recognized can sex be understood as requiring women's desire and pleasure. Currently, "its absence [is] acceptably unquestioned and unnoticed."[19]

Cahill mentions several ways in which empirical research bears upon her claims. Rosemary Basson posits that for many women, sexual desire is responsive rather than spontaneous.[20] There is an intuition to understand desire as freely chosen, say, through women's sexual fantasies and thoughts, rather than as an automatic, bodily arousal to sexual stimulus.[21] According to Basson, women are more likely than men to experience (often without knowledge) genital arousal with no simultaneous feelings of arousal or desire.[22] Most notable is women's sexual arousal responding to scenes of sexual violence, whether wanted or not;

14. Cahill, "Recognition," 307.
15. Kukla, "That's What She Said," 72.
16. Cahill, "Recognition," 310.
17. Cahill, "Recognition," 310.
18. Cahill, "Recognition," 310.
19. Cahill, "Recognition," 315.
20. Cahill, "Recognition," 311.
21. Cahill, "Recognition," 311.
22. Rosemary Basson, "The Female Sexual Response: A Different Model," *Journal of Sex Marital Therapy* 26, no. 1 (2000): 51-65.

some women also experience physical responses during incidents of rape or assault. Thus, spontaneous desire seems unachievable and difficult to disentangle from physical responses.[23] Inadequate understanding of women's experience of desire threatens to perpetuate a masculinist framing of desire instead of a gender-neutral one. With this in consideration, Cahill offers a model of desire that does not require desire to be spontaneous but encompasses all female desire as "embodied intersubjectivity."[24] That is, conceiving of desire as something sexual interactions can bring about, which is difficult to evaluate prior to any interaction. This then raises a series of questions: Can people have ethical sex if only one partner is desiring? Is it ethical to "elicit"[25] desire from a non (or not yet) desiring partner?

Cahill believes that it is possible to have ethical interactions with only one partner having sexual desire initially, and then engaging with the other intending to "create mutually felt desire."[26] She uses Gavey's example to argue that while it is not necessary to have a standard that requires both partners to physically want sex, the person who does not feel physical desire but chooses to give sex as a gesture of love or generosity should do so freely.[27] Desiring to give sex or desiring the physical arousal from sex are both ethical situations. For Cahill, the key question is not whether women desire sex, but that women's desire matter.

While this framing is certainly more robust than consent, I have two worries about desire as an alternative. First, much of desire is based on mutual wanting, which seems to be oversimplified in Cahill's model and to defeat some aspect of women's sexual subjectivity. Second, Cahill assumes that with desire, sex is ethical but does not question the origins of those desires, and whether certain desires (e.g., race play in BDSM) are inherently unethical. Beginning with the first concern, a significant component of desire stems from wanting to be wanted. While Cahill acknowledges this second order wanting in "embodied intersubjectivity," her framing of female agency seems to lose some traction as women often internalize others' desire as a sign of validation or self-worth. For example, the fact that a woman does not have control over whether a man finds her attractive makes his desire for her all the more validating or "genuine." Similarly, in "Cat Person," Margot desired sex with Robert because she knew that he found her attractive: "He looked stunned and stupid with pleasure, like a milk-drunk baby, and she thought that maybe this was what she loved most about sex—a guy revealed like that."[28] His desire for her made her more aroused: "the more she imagined his arousal, the more turned-on she got."[29] Margot's arousal certainly seemed dependent on Robert's desire, which then

23. Basson, "The Female Sexual Response," 51-65.
24. Cahill, "Recognition," 313.
25. Cahill, "Recognition," 313.
26. Cahill, "Recognition," 313.
27. Cahill, "Recognition," 313.
28. Roupenian, "Cat Person."
29. Roupenian, "Cat Person."

raises the question of whether it matters if desire is *of* the partner or from feeling wanted. If it is the latter, as the case with Robert, Cahill's desire model still is inadequate to prevent questionable sex. One can easily imagine similar scenarios where the woman engages in undesired sex while having an alternate fantasy of better sex to escape from reality, or even engaging just to feel desired. While Cahill encompasses this under her model, second-order desire is not equivalent to first-order and certainly elicits questions of what desire ought to be.

Furthermore, second-order desire can lead to harmful situations that further perpetuate a male-dominated sexual framing. For example, for many young girls or women who are not yet aware that their sexual subjectivity matters, much of their desire is dependent upon a man's. In a similar way to depending on makeup to feel pretty, a woman's self-perception could also arise from her male peers' desire; there is a certain type of empowerment or confidence she feels when she is wanted. Naturally, the woman gives up some sexual autonomy when she seeks validation from second-order desire. Of course, one could picture scenarios where both partners experience second-order desire equally, in which case mutual desire leads to intensified and positive sexual experiences. Yet, that would seem to depend on context, and second-order desire alone can lead to enduring problematic sex, which defeats some female subjectivity.

The ethics of eliciting desire also challenges Cahill's framing. While she acknowledges that it can be ethical to elicit desire from a partner, I worry about existing power dynamics that result in Gavey's gray cases, in which men are much more likely to elicit desire from women. In those cases, women are more likely to "give" sex (whether due to love, care or responsibility) than men. If there's no requirement for women to feel desire autonomously, the terrain seems slippery. Men could initiate unwanted sexual acts to elicit "responsiveness" from the woman, which may result in classic rape narratives of "You didn't know you wanted this." Of course, this ethical standard depends greatly on context and differs for lovers who are cheating on respective spouses (Gavey's stories), casual sex, or couples in long-term, committed relationships. Despite different contexts and understandings of desire, there should be some robust requirement of a woman's desire that is uniquely hers, regardless of how that takes form.

Lastly, Cahill takes it for granted that any sexual desire is ethical. However, this ignores potentially problematic desires that may bring pleasure, but raise ethical concerns. For example, Georg Friedrich Haas, a Columbia music professor, and his wife, Mollena Williams, frequently engage in race plays of domination and submission. While mutually consensual, their situation certainly raises eyebrows when Haas desires "to tame" Williams through master-slave race play.[30] While

30. Zachary Woolfe, "A Composer and His Wife: Creativity Through Kink," *New York Times*, February 23, 2016, https://www.nytimes.com/2016/02/24/arts/music/a-composer-and-his-wife-creativity-through-kink.html.

it may be pleasurable for both parties, the acts raise moral issues since they are reminiscent of the horrific legacy of slavery, which many people are not comfortable with sexualizing. While desire has largely been relegated as a realm free from moral evaluation, it is highly plausible that people's desires are formed from problematic social forces. As such, desire is not independent of outside influences and cannot be neatly set aside in one's sexual life. While desires are not subjected to rational or moral evaluation in the same way that beliefs are, there should be moral requirements on people's desires apart from consent. If this were true, Cahill's advocacy of desire would seem to need some defense even from itself. While that requirement may be ambiguous, Cahill has not adequately addressed the harms of second-order desire and desires that stem from problematic beliefs. In considering other alternatives, it is now helpful to turn to Kukla, who conceives of consent as a speech act.

IV. Sexual Negotiation as an Alternative

Kukla argues that a language of sexual negotiation would suffice as an alternative to consent. In the current discourse, sexual communication is caught in a bind: clear and direct communication is important, yet much of erotic language is indirect.[31] As Kukla notes, "the language of flirtation, seduction, and engagement . . . tends to be circuitous, stagy, elliptical, metaphorical, innuendo filled, and connotative. This is part of what makes it sexy."[32] Yet, the ambiguity around erotic communication, whether physical or verbal, has the potential to be tremendously harmful for those involved and, in extreme cases, may lead to rape. Kukla aims to create "an alternative discursive frame" that revolves around sexual negotiation, which bridges the tension between clear and elusive communication.[33] This dialogue focusses on body language and connotative norms; for example, people do not take "Oh, tear me open!" literally and interpret the butt lift as a woman's invitation to remove her underwear.[34]

Kukla writes that role-playing or BDSM are good examples of when nonliteral speech is most explicit. While popular discourse offers slogans such as "no means no," part of the eroticization in BDSM is to flip normative meanings and obscure consent.[35] For example, safe words replace the function of 'no,' which would be taken to mean anything but stop in BDSM. To achieve an ethical standard in sex, partners must develop a sexual language to understand when 'no' means stop and when 'no' elicits further eroticization.

Importantly, Kukla points out that sexual encounters can easily slip in and out of comfort zones, so that literal speech may not suffice. For example, women often do not feel like they have the language to even refuse or convey what they want.

31. Kukla, "That's What She Said," 79.
32. Kukla, "That's What She Said," 79.
33. Kukla, "That's What She Said," 79.
34. Kukla, "That's What She Said," 79.
35. Kukla, "That's What She Said," 80.

Rather, people must learn to understand gestures, nonliteral speech and feelings as part of consent and successful sexual communication. A robust negotiation model extends beyond consensual acts, when one refuses or accepts an action to be done.[36] The negotiation model encompasses a range of discourse and active participation "in deciding not just whether to have sex but what to do," encouraging partners to share desires and maximize pleasure for all involved.[37] Practically, the model would serve as "mutual consultation," considering each person's tastes and assuming reciprocal responsibilities.[38] Ideally, this results in egalitarian sex where both parties are equal in sexual subjectivity.

Note that the negotiation model need not oppose Cahill's framework of desire, and in some ways, Kukla's model is a powerful addition in ensuring both parties involved have full autonomy in cases with the absence of desire. While Kukla's model is compelling, I am concerned with its applicability to casual sex and in practically countering women's self-regulation.

V.

Kukla's language of sexual negotiation does little for casual sex, where partners are not familiar enough to understand each other successfully and ethically. Yielding sexual control to a trusted partner and devising elaborate safe word systems in role-playing seems to differ greatly from one night stands. It would be unrealistic to require the same uptake and interpretation in long term relationships as in casual sex. While BDSM negotiations offer the communication system that Kukla had envisioned, it is difficult to imagine sexually budding teenagers using those systems or expecting these "felicity conditions" to be in place if the partners are yet to be made aware of them. In these cases, consent seems to be a better evaluation method for just sex, despite its shortfalls. Casual sex, however, raises an interesting question about how much explicit communication is actually needed contrary to Kukla's nonliteral speech. When there is little familiarity with a partner's body language or preferences, nonelusive speech seems most effective in avoiding the dangers of miscommunication. Regardless, negotiation requires women to have a type of courage to assert themselves sexually that Kukla takes for granted.

Gavey's gray cases are a good starting point for where Kukla's model may fall short. In those situations, women are portrayed as passively giving into sex, often with people that they know. Even Kukla acknowledges that men often misunderstand women's nonverbal behavior, and "impute erotic innuendo and sexual intent where there is none."[38] While the negotiation model requires a conversation that involves invitations and exchanges, these women are unlikely

36. Kukla, "That's What She Said," 93.
37. Kukla, "That's What She Said," 93.
38. Kukla, "That's What She Said," 93.
38. Kukla, "That's What She Said," 93.

to take pragmatic actions to counter the heterosexual coercive forces to which they fall prey. If they cannot refuse in the minimal consent model, it is unrealistic to expect Gavey's interviewees to challenge their partners and negotiate on their own terms. If the expectation is that like in BDSM, negotiation occurs before sex, there is always a danger of leaving something out or for some, not even knowing what their comfort zones are until they have been violated. Even if women in gray cases were to negotiate, one could easily imagine one-sided negotiations in which the woman is compromised because she feels compelled to give more rather than receive. The resulting sex may even be worse for women, as their partner may feel that anything within the terms of negotiation is fair game. Lastly, if fair negotiation cannot happen, Kukla's model would unnecessarily burden women to strive towards negotiation when they have yet to feel enough sexual agency to refuse; in this case, the responsibility would further be placed on women who failed to negotiate fair terms for themselves. As such, Kukla is introducing a radical model that is highly desirable, but unrealistic considering current societal forces that must first be addressed.

From here, one may be wondering what a positivist view of just sex looks like. I propose a mixed framework that involves consent, desire and negotiation. All three factors are important, but they are to be weighed differently depending on the gender, orientation, tastes and familiarity of the parties involved. In effect, the framework for evaluating just sex in one-night stands is different than in long term relationships. For example, without the necessary conditions for good communication in casual contexts, or as a matter of fact, not much information about what the parties involved like and dislike, consent is weighed much more heavily as a requisite. Note here that the casual sex in "Cat Person" would not qualify as fully consensual due to the discomfort Margot experienced following what began as a seemingly harmless situation. On the other hand, in contexts where the parties involved are very familiar with each other in a healthy, long-term relationship, negotiation that includes mutual desire is more suitable than merely consenting to sex.

VI. Conclusion

A key takeaway from this paper should not be dismissing the alternatives of desire and sexual negotiation due to their respective shortfalls, but rather, to envision them in a broader picture of prioritizing women's sexual agency that buttress the bare bones of consent. Desire and negotiation are parts of a whole that consist of women's inter-subjectivity in sexual encounters. While there may not be a satisfactory one size fits all requirement for ethical sex, the general consensus is that consent alone is not enough, and we need more rigorous standards. Gavey's gray cases establish the slippery terrain of sex as something women feel little control over, as consensual beginnings can easily transition into unwanted sex that elicits deep discomfort, harm and shame. To address gray sex, Cahill offers the desire framework that poses women as inter-subjective agents able to experience desire however they wish. Yet, the worry here is that much of

desire is dependent on mutual want, which undermines some of women's subjectivity; and second, that some desires may be inherently problematic. Kukla brings forth the negotiation model that enables both partners to equally decide the terrain for sex, beyond assenting or dissenting. Yet, this model has little applicability in casual sex when there is no familiarity to establish a robust, mutual sexual language; and second, it is difficult to imagine women negotiating fair terms for themselves in this model when they have yet to counter the coercive forces of self-regulation. Therefore, neither model is sufficient on its own for ethical sex, but a mixture of all three, consent, desire and negotiation, depending on situation, adds up to a viable illustration of what just sex should be.

Bibliography

Cahill, Ann J. "Recognition, Desire, and Unjust Sex." *Hypatia* 29, no. 2 (2014): 303–19.

Gavey, Nicola. "Technologies and Effects of Heterosexual Coercion." *Feminism & Psychology* 2, no. 3 (October 1992): 325-351.

Kukla, Rebecca. "That's What She Said: The Language of Sexual Negotiation." *Ethics* 129, no. 1 (October 1, 2018): 70–97.

Roupenian, Kristen. "Cat Person." *New Yorker*, December 4, 2017. https://www.newyorker.com/magazine/2017/12/11/cat-person.

Woolfe, Zachary. "A Composer and His Wife: Creativity Through Kink." *New York Times*, February 23, 2016. https://www.nytimes.com/2016/02/24/arts/music/a-composer-and-his-wife-creativity-through-kink.html.

The UBC Journal of Philosophical Enquiries