

Northwestern University Society for the Theory and Ethics of Politics

-12th Annual Conference-

March 8th - 10th, 2018

John Evans Alumni Center 1800 Sheridan Rd., Evanston



Keynote Addresses:

“Official Corruption”

-Niko Kolodny, UC Berkeley-

and

“Agency and Vulnerability”

-Sharon Street, NYU-

A special thank you to The Weinberg College of Arts and Sciences for its
generosity

Table of Contents

SCHEDULE.....	2
Acknowledgements.....	4
Keynote Address: <i>Agency and Vulnerability</i>	5
Sharon Street	
Keynote Address: <i>Official Corruption</i>	19
Niko Kolodny	
<i>Desire and Loving for Properties</i>	45
Yongming Han	
<i>Mutual Recognition, Friendship, and the Moral Relationship</i>	55
Richard Healey	
<i>Suffering, Moral Luck, and the Duty to Self-Care</i> Alycia LaGuardia-LoBianco	
<i>Grace and Alienation</i>	71
Vida Yao	
<i>From Rationality to Morality: The Collective Development of Practical Reason in Kant's Anthropological Writings</i>	85
Olga Lenczewska	
<i>On Being a Rational Animal and Public Reasons</i>	99
Troy Seagraves	
<i>Everywhere Chimerical</i>	108
Jorah Dannenberg	
<i>Solidarity and Community</i>	120
Mike Zhao	
<i>What's Wrong with Gentrification?</i>	131
Rafeeq Hasan	
<i>Pitting People Against Each Other</i>	142
Waheed Hussain	
Chicago Attractions	168
Map of Downtown Evanston	169

Northwestern University Society for the Theory of Ethics and Politics

12th Annual Conference
March 8-10, 2018
John Evans Alumni Center

Thursday, March 8th

9:00-10:25

“Desire and Loving for Properties”

Yongming Han

Commenting: Regina Hurley

10:35-12:00

“Mutual Recognition, Friendship, and the Moral Relationship”

Richard Healey

Commentator: Aleksy Tarasenko-Struc

Lunch

2:15-3:40.

“Suffering, Moral Luck, and the Duty to Self-Care”

Alycia LaGuardia-LoBianco

Commenting: Susan Bencomo

3:50-5:15

“Grace and Alienation”

Vida Yao

Commentator: Laura Gillespie

Dinner

Friday, March 9th

9:00-10:25

“From Rationality to Morality: The Collective Development of Practical Reason in Kant's Anthropological Writings”

Olga Lenczewska

Commenting: Hao Liang

10:35-12:00

“On Being a Rational Animal and Public Reasons”

Troy Seagraves

Commenting: Paymun Zargar

Lunch

2:15-3:40

“Everywhere Chimerical”

Jorah Dannenberg

Commentator: Ryan Davis

3:50-5:45.

Keynote Address:

“Agency and Vulnerability”

Sharon Street

Commentator: Andrea Westlund

Reception – Everyone is invited

Saturday, March 10th

9:00-10:25

“Solidarity and Community”

Mike Zhao

Commenting: Eskil Elling

10:35-12:00

“What's Wrong with Gentrification?”

Rafeeq Hasan

Commentator: Paul Schofield

Lunch

2:15-3:40.

“Pitting People Against Each Other”

Waheed Hussain

Commentator: Oded Na'aman

3:50-5:45

Keynote Address:

“Official Corruption”

Niko Kolodny

Commentator: Paul Weithman

Dinner

Acknowledgements

Conference Organizers:

Kyla Ebels-Duggan, Richard Kraut, Stephen White, and Abby Bruxvoort

For organizational and administrative assistance, special thanks to Susan Bencomo, Josh Kissel, Hao Liang, Carmen De Schryver, Michael Swartz, Crystal Foster, Ozge Hemmat, and Ben Miller.

Special thanks to the Weinberg College of Arts and Sciences for its generous support

Note to 2018 NUSTEP conference participants: Below is an outline of the talk I'll be giving. This material is part of a larger book project, tentatively titled *A Secular Foundation for Ethics: Attachment, Loss, and the Standpoint of Pure Awareness*. The aim of the book is to sketch out a metaethical view that is metaphysically antirealist, and yet which nevertheless vindicates a strong form of ethical objectivity—in part by drawing upon insights that I believe may be found in eastern meditative traditions. Roughly, the idea is to start by identifying what I call the *standpoint of pure awareness*, a “way of looking” that may be characterized, or at least pointed to in an informative way, by appealing to the form of attention cultivated in meditation practice. The next step to argue that the standpoint of pure awareness has the following three features: (1) it offers a solution to a deep problem that all of us face in virtue of being agents; (2) it involves an accurate seeing of an underlying metaphysical reality (roughly, the lack of difference between “self” and “other”); and (3) it constitutes an ethical perspective. In this talk, I'm focusing on one very limited component of this larger argument, namely the task of identifying the “deep problem” that I claim all of us face in virtue of being agents. I've already taken one stab at this task in a recent paper, “Constructivism in Ethics and the Problem of Attachment and Loss” (*Proceedings of the Aristotelian Society, Supplementary Volume* (2016): 161-189), but I don't think that paper succeeded in locating the problem clearly enough. So in this draft I'm working to try to nail down the problem more precisely. However, as is obvious from the outline format, this is very much work in progress, so please don't cite or quote without permission.

Agency and Vulnerability

Sharon Street
New York University
February 19, 2018

1. Introduction

- a. The “Holy Grail” of secular analytic metaethics = a philosophical account of the subject matter of ethics that both (1) vindicates a strong form of ethical objectivity and (2) does so without metaphysical and epistemological mystery.
- b. I am the first to be skeptical of the existence of this “Holy Grail,” having elsewhere defended a rather strongly subjectivist view.¹ Still, I think it would be premature to give up the search, for two reasons (among others):
 - i. First, I'm convinced that we—the “we” here being those of us who work in secular analytic metaethics in the western tradition—have not even begun to exhaust the available philosophical resources, and that a look to eastern meditative traditions, including Advaita Vedanta and Buddhism, shows this.

¹ I describe and defend the subjectivist implications of my view in “Coming to Terms with Contingency: Humean Constructivism about Practical Reason,” in *Constructivism in Practical Philosophy*, eds. James Lenman and Yonatan Shemmer (Oxford: Oxford University Press, 2012), pp. 40-59; “In Defense of Future Tuesday Indifference: Ideally Coherent Eccentrics and the Contingency of What Matters,” *Philosophical Issues*, vol. 19 on “Metaethics,” ed. Ernest Sosa, 2009, pp. 273-298; and “Nothing Really Matters, But That's Not What Matters,” in *Does Anything Really Matter? Parfit on Objectivity*, ed. Peter Singer (Oxford, Oxford University Press, 2017), pp. 121-148.

- ii. Second, the vindication of the objectivity of ethics is not an all-or-nothing affair. We may learn a lot by aiming high and seeing where, short of that, we land.
- c. In my view, our best hope for vindicating a strong form of ethical objectivity, while avoiding metaphysical and epistemological mystery, is *metaethical constructivism*, a view that I've argued for elsewhere.²
- d. Glossing over many details,³ we may characterize the *generic metaethical constructivist strategy* for vindicating the objectivity of ethics as consisting of three steps:
 - i. **Step 1:** Explain what is constitutively involved in being an agent, valuer, or will at all.⁴
 - ii. **Step 2:** Identify a universal problem that is faced by every agent, by her own lights, simply in virtue of her being an agent at all.
 - iii. **Step 3:** Argue that the (best, or only?) solution to the relevant problem involves adopting an ethical principle or standpoint.
- e. Korsgaard's Kantian implementation of the strategy⁵:
 - i. **Step 1:** Explain what is constitutively involved in being a *will*.
 - ii. **Step 2:** Argue that every will has the problem of "needing a law" that is not "given from the outside."

² See "Constructivism about Reasons," *Oxford Studies in Metaethics*, vol. 3, ed. Russ Shafer-Landau (Oxford: Clarendon Press, 2008), pp. 207-245; and "What is Constructivism in Ethics and Metaethics?" *Philosophy Compass* 5 (2010): 363-384.

³ For more details about what I'm calling the "generic" metaethical constructivist strategy, see "Constructivism in Ethics and the Problem of Attachment and Loss," *Proceedings of the Aristotelian Society, Supplementary Volume* (2016): 161-189.

⁴ Different implementations of the generic constructivist strategy may make different choices about the most useful basic concept to start with—whether it is *agent*, *valuer*, *will*, or something else. Regardless of the concept chosen, Step 1, properly conducted, involves purely "conceptual" work in the sense that it just involves sketching the kinds of things that a being needs to be doing in order recognizably to qualify as an agent (or valuer or will) at all. A crucial point that must be respected in this conceptual work is the point that an extraordinarily wide range of "values" is compatible with someone's recognizably being an agent (or valuer or will). Most relevantly for our purposes, grossly immoral values are perfectly compatible, as a conceptual matter, with recognizably being an agent (or a valuer, or having a will). To smuggle moral values into one's account of what is conceptually involved in being an agent, as some theorists more or less obviously do, makes the resultant account worthless as a potential vindication of the objectivity of ethics.

⁵ See especially her "Realism and Constructivism in Twentieth-Century Moral Philosophy," *Journal of Philosophical Research*, APA Centennial Supplement (2003): 99-122.

- iii. **Step 3:** Argue that the solution to that problem is the categorical imperative as represented by Kant's Formula of Universal Law.
- f. An alternative implementation of the strategy that I am exploring in current work:
 - i. **Step 1:** Explain what is constitutively involved in being a valuer.⁶
 - ii. **Step 2:** Argue that no matter what the particular substantive content of a given agent's evaluative perspective, she faces the problem of vulnerability to devastating or total loss.
 - iii. **Step 3:** Argue that the (best?) available solution to the this problem is to seek out and occupy what I call the *standpoint of pure awareness*. Then argue that the standpoint of pure awareness involves both (1) an accurate seeing of reality and (2) taking up an ethical perspective. I will not be discussing this step at all in this paper, but the idea is to draw on insights that I believe may be found in eastern meditative traditions, especially Advaita Vedanta and Buddhism.⁷
- g. In this paper, I will be focusing exclusively on Step 2, hoping to arrive at a clear statement of the relevant problem, and taking a stab at arguing for the maximally ambitious claim that the problem in question is universal to agency. While I think it's worth trying for this most ambitious claim and seeing where the argument leads, I'll be acknowledging that there is not an airtight connection here. Ultimately the fallback position is simply that among agents like us (in other words, human beings), the problem is universal or nearly so. That still is an ambitious claim that one might question.

⁶ I say more about this step in "Constructivism in Ethics and the Problem of Attachment and Loss" and elsewhere. A very brief summary of some relevant points: First, I use the term *valuing* as a term of art meant to encompass a wide range of attitudes and experiential states. Second, in seeking to characterize this family of attitudes, what will *not* work, in my view, is to try to give a reductive characterization. On the contrary, there is no way accurately to characterize the relevant attitudes without invoking normative concepts (e.g., saying things like they're the attitude of taking something to be *good* or *required* and so forth). Third, the fact that we cannot accurately characterize the attitudes in question without invoking normative concepts doesn't mean we can't say anything informative about them and locate them perfectly well for ourselves. For one thing, we can get a fix on the relevant *concepts* by pointing to circumstances in which we tend to experience things as *good*, *valuable*, *worthwhile*, *called for*, etc. (Cf. color concepts.) We may furthermore characterize the relevant set of attitudes by contrasting them with the attitude of *mere desire*. Elsewhere (in "Coming to Terms with Contingency") I discuss three points of contrast. Lastly, we may point to loose constitutive connections with emotions such as anticipation, hope, joy, dread, relief, outrage, anxiety, fear, sadness, and regret. The resulting account of the attitude of valuing will be loose and imprecise, but this is not a problem for two reasons. First, it gets the phenomenon right: for the phenomenon in question isn't a simple and perfectly unified phenomenon; intuitively, we see a wide range of forms of "valuing" in the animal world, and the account captures this. Second, a loose and imprecise characterization of the attitude is all that's necessary for the metaethical project we are pursuing.

⁷ I say some sketchy things about this step in "Constructivism in Ethics and the Problem of Attachment and Loss."

2. Two clarifications of the task

- a. Before starting in with my attempt to implement Step 2, let me offer two clarifications of what we're in search of here.
- b. First, note that it is consistent with an agent's "having a problem" in the relevant sense that she does not currently (and indeed may never) recognize the problem as hers. The idea is rather that there is a problem that every agent does indeed have, by her own evaluative lights, when these evaluative lights are combined with the non-normative facts. But she might not now (or ever) recognize or agree that she has that problem. This could happen if she is ignorant of, or incapable of recognizing, relevant non-normative facts, or if she is for whatever reason unable to see or draw out the logical and instrumental implications of her deepest values.
- c. Second, note that it is consistent with an agent's "having a problem" in the relevant sense that he already has, in his possession, a solution to it—maybe a perfect and complete solution, or maybe just a makeshift and partial one. For example, there is a sense in which it makes sense to say that all of us here in this room have the problem of obtaining food, water, shelter, and clothing. It's an ongoing challenge we face in virtue of pretty deep (in this instance, biological) features of ourselves. As it happens, those of us in this room are fortunate enough to be currently in possession of an ongoing solution to these problems. The larger point: When it comes to the position I want to defend, it is consistent with there existing a certain problem, faced universally by all agents in virtue of their being agents at all, that many agents already have the problem under control. After all part of the ultimate point, if the overarching position I want to suggest pans out, is supposed to be that there *is* a solution to the problem in question, and that it involves taking up an ethical perspective.
- d. I'll now start working to try to pinpoint the problem I have in mind, beginning with a rough, intuitive statement. I'll then work to develop the account in stages—first noting a constitutive connection between valuing and vulnerability, and then moving through a series of four refinements on the way to a sharper statement of the problem that I want to claim all agents face.

3. The rough, intuitive idea

- a. We are small, finite beings in a very big universe that does not, in itself, care a whit about how things turn out for us.⁸ And yet, as valuers, we are invested in how things turn out—in what happens to ourselves and the people and things we love—and this makes us deeply vulnerable. Indeed, we are *so* vulnerable

⁸ It is a background assumption of this paper that God, in the sense of a perfectly good, omniscient, and omnipotent agent, does not exist. I defend this assumption (putting forward a version of the argument from evil) in "If 'Everything Happens for a Reason,' then We Don't Know What Reasons Are: Why the Price of Theism is Normative Skepticism," in *Challenges to Religious and Moral Belief: Disagreement and Evolution*, eds. Michael Bergmann and Patrick Kain (Oxford: Oxford University Press, 2014), pp. 172-192.

that a clear-sighted awareness of that vulnerability can have a corrosive effect on our ability to go forward, leading to crippling anxiety or depression, for example. Our hopes and loves, conjoined with our finitude, make us vulnerable to loss—indeed, in many cases guarantee it—and yet we don't want to abandon our hopes and loves. How best to cope with this situation? Is there a way of coping that doesn't involve some version or other of denial, distraction, numbing out, or hardening? How should we live with ever-present threat (and indeed guarantee, for those of us who never want to die or grow old or be separated from the ones we love) of losses that we reasonably fear we might not be able to withstand?

- b. Cf. the character Nachiketa's dialogue with Yama, the King of Death, in the *Katha Upanishad*. The King of Death offers Nachiketa wealth and power, a long life, children and grandchildren who will themselves live "a hundred years," and a time on earth full of worldly delights. Nachiketa replies to Death:

These pleasures last but until tomorrow,
And they wear out the vital powers of life.
How fleeting is all life on earth! Therefore
Keep your horses and chariots, dancing
And music, for yourself. Never can mortals
Be made happy by wealth. How can we be
Desirous of wealth when we see your face
And know we cannot live while you are here?
... How can I, subject to old age and death,
Ever try to rejoice in a long life
For the sake of the senses' fleeting pleasures?⁹

- c. Nachiketa is in the grip of a problem and can't focus on anything else until he finds a solution or dies trying. Acutely aware of the eventual loss of everything that he loves and enjoys, he finds himself without the will to pursue any of it. Other things being equal, the things Death offers him are good: a long life, the promise of thriving children and grandchildren, influence and respect, enjoyment of the arts and creature comforts, etc. But

⁹ *The Upanishads*, introduced and translated by Eknath Easwaran (Tomales, California: Nilgiri Press, 2007), p. 74.

Nachiketa finds his interest in these things gutted by his awareness that one day they will all be lost.

- d. Nachiketa's concerns here are very human, yet one might think the basic structure of his problem is universal—dependent only on the fact that he is a finite valuer. I'll now try to develop that thought.

4. First stab at a philosophical account: Valuing and vulnerability

- a. Start with the following observation. However exactly we characterize the attitude of valuing (a task I say more about elsewhere), it seems plausible that constitutively involved in the attitude of valuing is taking there to be either an actual gap, or at least the possibility of a gap, between the way the world as it as a matter of fact *is*, on the one hand, and the way it would be *good* or *desirable* for the world to be, on the other. This is not intended to be a substantive point, but merely a conceptual observation.
- b. If this is so, then (one) problem that is necessarily built into the standpoint of any valuer, as such, is the problem of a gap, or the potential for a gap, between the world as it as a matter of fact is, and the world as it would be good or desirable (according to that valuer's standpoint) for the world to be.
- a. Now couching the same point in somewhat different terms, to be a valuer is (among other things) to be a being who is attached to the world's being one way rather than another, and who, in virtue of that attachment, is vulnerable to the way in which the world unfolds. To be a valuer is to be a being for whom there is always at least a potential divergence between how things are and how (according to her evaluative perspective) it would be good for things to be, such that the valuer is subject to the ever-present threat (or, depending on the precise content of her values, guarantee) of loss.
- b. Note that I am using the terms *attachment* and *loss* in a very general way. We may distinguish two senses of *attachment*:
 - i. *Attachment in a narrow sense* = the sense in which an "attachment" is something that can only exist with respect to an already or previously possessed object, person, place, or time.
 - ii. *Attachment in a broad sense* = the sense in which one can be "attached" to things that are not yet (and perhaps never will be) present, actualized, or possessed.
- c. We may distinguish two corresponding senses of *loss*:
 - i. *Loss in the narrow sense* = the sense that requires having *had* something good in order to be able to lose it.

- ii. *Loss in a broad sense* = the sense of loss according to which one need not ever have possessed something in order to suffer a loss; on this broader understanding, failure, disappointment, and heartbreak over things never achieved, possessed or realized count as losses.
 - d. Using the terms *attachment* and *loss* in their broadest senses, then: To be a valuer is to be attached to things (people, places, objects, ways of life, certain states of the body and mind, the world's unfolding one way rather than another, and so on) in a way that makes one vulnerable to loss.
 - e. We're far from having pinpointed the universal problem of agency that I ultimately want to isolate, but it is a start.
5. First refinement: The possibility of invulnerable valuers
- a. My remarks so far might seem to suggest that every conceivable valuer faces the problem in question.
 - b. But this is not so. To see this, we need to consider the possibility of what I'll call *invulnerable valuers*.
 - c. There are at least two types of invulnerable valuers we might imagine:
 - i. *Valuers who are invulnerable in virtue of being omnipotent*. I won't spend any time on this possibility, because I take it to be clear that the universe does not contain omnipotent valuers. Individual, conscious evaluative points of view on the world arise at particular places in space and time, due to particular sorts of configurations of matter, and those configurations are continually changing and impermanent. Any being with a conscious evaluative standpoint on the world depends for its existence on a particular ongoing organization of matter, which will eventually move out of existence. In more everyday terms, valuers are finite; we are subject to illness, accident, aging, death, and more generally circumstances beyond our control.
 - ii. *Valuers who are invulnerable in virtue of the nature of their ends*. A far-fetched example useful for the sake of illustration: a valuer who values the fact that $2 + 2 = 4$. More realistically: a valuer who values the fact that he already accomplished something (e.g., wrote the great American novel)—a fact that the rest of time can do nothing to change.
 - d. Let's now take a closer look at this second type of invulnerable valuer.
 - e. First off, notice that in order to be this type of invulnerable valuer, it is not enough that one values the above sorts of things; it must also be the case that this is *all* one values, and such cases are hard to imagine.

- f. Nevertheless, it seems to me that we *can* imagine such valuers. They wouldn't live very long, since they wouldn't be motivated to pursue the means necessary for living or doing anything, but we can still imagine them.
 - g. If this is right, then it's not true that every *conceivable* valuer faces the problem of attachment and loss. The possibility of invulnerable valuers shows that we've failed to find a universal problem facing *valuers* as such.
 - h. But the next question is whether this puts a damper on the larger philosophical aim that we are pursuing—namely seeking to vindicate the objectivity of ethics by way of the generic metaethical constructivist strategy. My claim is that it does not.
 - i. For notice that invulnerable valuers have *no reasons for action*.¹⁰ They might be agents in the sense that they are *capable* of acting for normative reasons, but due to their happy combination of values and the actual state of the world, there is nothing at all such that they have normative reason to do it. The world, for them, is normatively perfect just as it is. So unless they are confused, these agents will not be in the business of action at all. If they act at all—e.g., eating and drinking each day so that they can perform a ritual that they falsely believe is required to keep the sum of 2 and 2 from going up to 5—then they are making a mistake. *A fortiori*, they are making a mistake if they act *unethically*.
 - j. So invulnerable valuers are not part of the audience that a vindication of ethics needs to address. They have no reasons to do anything at all. If they start to do *anything*—and *a fortiori* something unethical—they are making a mistake on their own terms (which, according to constructivism, are the only terms there are).
6. Second refinement: The problem isn't "vulnerability to loss" per se
- a. What I've said so far might make it sound like I think the problem in question is being vulnerable to loss *at all*—such that solution to the problem, if there is one, would involve becoming completely invulnerable to loss of any kind.
 - b. But I take it that for many or most of us, it's obvious that complete invulnerability to loss is not what we want. What would a life be with no risk, no vulnerability, no chance of failure, no challenges or ups and downs? While this point ultimately needs a fuller treatment, I take it that many people will readily agree with me that it's an immature perspective that regards eliminating *all* risk, *all* possibility of loss, as a worthy or desirable goal. After all, one could do *that* by opting out of valuing entirely, but that's not what most of us want to do. The widely shared intuition I have in mind is captured by the saying that it's better to have loved and lost than never to have loved at

¹⁰ Here (as throughout the paper) I am assuming an antirealist understanding of normative reasons.

all. It's also captured by the immediate negative reaction that many people have (I think mistakenly) to Buddhism or Stoicism upon first, superficial encounter with such views. On superficial inspection, both views can give the impression of recommending that we stop desiring or caring at all. While this is not the correct way to interpret these views, my point here is just that *if* that's the suggestion, it's one that many of us find totally unappealing. Loss is no fun, but the solution isn't not to play the game.¹¹

- c. So the problem is not vulnerability to loss in and of itself. Nor is the problem even, I think, that we're vulnerable to *extremely serious loss*. As maybe fewer but still many of us would agree, it's better to have loved with all one's heart, and to have lost terribly, than never to have loved at all.
- d. That being said, the prospect of loving with all one's heart and losing terribly—while perhaps better than never loving at all—is a frightening prospect. Indeed, it's a sickening prospect when contemplated from the standpoint of our deepest values and attachments. And it's a prospect that all of us face insofar as we do go ahead and love with all our hearts—for example, when we fall in love, or hope for a better future for humanity, or decide to have a child, even though we are finite beings living in a world that is radically beyond our own control.
- e. We're getting closer to the problem here: The real problem is one that becomes vivid, from a first-personal point of view, in the case of our deepest attachments—to our parents, children, and siblings; to our lovers and friends; to our youthful visions of the life we wanted to lead; to our hopes for a peaceful world; to our bodies and minds and life itself.
- f. How to put the problem? What bothers us, I take it, isn't vulnerability to loss, as such, which is part and parcel of being a valuer at all, but rather, as we might say, vulnerability to *devastating or total, unmitigated* loss.
- g. I take it to be a fact—easily established by observation—that the world routinely deals out losses of this kind. Skip the following sentences if you don't need convincing, but for anyone who does, consider: Grown men die in terror and agony, screaming for their mothers, in wartime. First-graders are shot to death in their elementary school classroom. Tsunamis sweep away tens of thousands of people within minutes. Less dramatically, but on countless occasions per day on planet earth, people's hearts are broken in love and friendship; their bodies and minds are destroyed by aging or illness; they realize that the life they always dreamed of is never to be; they're separated by

¹¹ [Add note on Epictetus: "You will never have to experience defeat if you avoid contests whose outcome is outside your control." One may be forgiven for getting the impression that Epictetus is recommending that one should therefore avoid such contests entirely. But it's clear from much else he says that this cannot be the suggestion.]

death from the ones they loved the most; or they die alone and in pain, believing no one cares about them.

- h. So let's try this: The problem in question is vulnerability to what I'll call, in a familiar turn of phrase, *devastating* loss. We're okay with vulnerability to loss, we might say; it's our *radical* vulnerability, our vulnerability to total, unmitigated loss, that is the problem. This is the kind of vulnerability awareness of which has the potential to eat away at our agency from the inside—crippling or sapping our ability to continue in pursuit of our ends at all.

7. Third refinement: Mellow and low-aiming valuers

- a. Putting the problem this way, however, opens the way for a different sort of objection, insofar as we're on the hunt for a genuinely universal problem of agency. While it was perhaps plausible to claim that all valuers, simply in virtue of being valuers at all, are vulnerable to loss of some sort or other (setting aside cases like the invulnerable valuers already discussed), it is far from obvious that all valuers, simply in virtue of being valuers at all, are vulnerable to *devastating* loss.
- b. In particular, there seem to be at least two ways a given valuer might be structured, such that she is vulnerable to loss, but not vulnerable to *devastating* loss:
 - i. First, there's the possibility of *extremely mellow valuers*: A valuer might simply be deeply easygoing—valuing plenty of things, and therefore vulnerable to loss, but with a low-key attitude that makes him relaxed even about the possibility of losing everything.
 - ii. Second, there's the possibility of *extremely low-aiming valuers*: A valuer might be vulnerable to *devastating* loss—roughly in the sense that she is deeply emotionally invested in certain outcomes, such that *if* the loss were to occur it would be devastating—but her hopes and ends might be *so* modest, *so* minimal, so exceptionally easy to achieve or see realized, that we might think she isn't in any meaningful sense genuinely vulnerable to the loss.
 - iii. And of course one might be both extremely mellow and extremely low-aiming at once.
- c. These seem like real possibilities. Indeed, they might even sound a bit like people we know. What to say about them?
- d. One point is that such valuers are not as common as one might think. For example, in the case of extremely low-aiming valuers, it is important to remember that while a given end on a particular occasion might well be trivial to the point of being fail-safe, what it will have taken to *get* to that point in the

creature who has reasons to act at all, and therefore not part of the audience that a vindication of ethics needs to address.

- j. Thus, although I've made a significant concession, I think there is still a deep constitutive connection, albeit it not airtight, between valuing and the problem of vulnerability to devastating loss. The more *strongly* a being values things, and the more high-aiming she is, the more she is subject to the problem of vulnerability to devastating loss. And while it's true that the mellow and more low-aiming she is, the less vulnerable she is to devastating loss, it's also true that she is fading out of the game of agency at all.
- k. A final point: Even a very mellow or low-aiming valuer is subject to *total* loss, in the sense of total failure on his own terms, even if those terms are very modest and relaxed. Insofar as that is true, he is also subject to whatever anxiety his nature can muster on behalf of any unlikely, low-stakes outcome. To that extent, even *he* may have reason to be interested in a solution for coping with the possibility of total loss.

8. Fourth refinement: The problem isn't suffering

- a. As I've mentioned, one of the aims of the larger project I'm pursuing is to draw upon eastern meditative traditions to try to make progress in secular analytic metaethics, and there are obvious parallels between the problem I've been pointing to and problems emphasized in the Buddhist tradition, which similarly focuses on how our desires combine with our finitude to generate a problem.
- b. There is a great deal from the Buddhist tradition that I ultimately wish to take up in one form or another, but there is an important difference that I want to highlight between the position I'm trying to develop and the Buddhist position, at least as that position is most commonly articulated and understood.
- c. As is often pointed out, Buddhism is a deeply practically-oriented view—seeking to pinpoint, diagnose, and then solve a very deep problem that all human beings face. This makes it potentially very well suited for marriage with the metaethical constructivist strategy for vindicating the objectivity of ethics. Furthermore, the lead seems especially promising given that there is (according to the Buddhist position) supposed to be a deep and indeed even necessary link between (1) the proposed solution to the problem and (2) an attitude of compassion for all sentient beings.
- d. I think the lead is indeed promising, but at the same time I believe that adjustments to the Buddhist view, at least as often characterized, are called for. And one such adjustment is in our characterization of the *problem*. According to the Buddhist position (at least as most often expressed), the most

fundamental problem faced by human beings is *suffering* (or *dukkhā*¹²). The First Noble Truth states that suffering exists; the next three seek to diagnose the source of suffering and to suggest a path (the Eightfold Path) for escaping it.

- e. In my view, to frame the problem this way is to get things just a little bit off from the beginning. To cast *suffering* as the main problem is to misdescribe what the problem really is from the relevant standpoint, which is *our* standpoint—the standpoint of living (suffering!) agents. The problem, as it presents to us, is first and foremost the gap between what *is* and what would be *good*. Suffering is what happens when that gap gets to be too large, too hopeless, too seemingly beyond repair; but the suffering is normally the manifestation of our taking something *else*—not the suffering itself—to be the problem. The problem, the mother who has lost her child wants to say, is not that I’m suffering: the problem is that my child is dead and never coming back. I’d gladly suffer for a thousand years if you could just bring him back. *He* is what matters, she’ll say, not my state of mind.¹³
- f. There is a sense in which this adjustment is cosmetic, but another sense in which it is absolutely not. To put the focus on what is most naturally understood as an *experiential state* is to build in from the beginning a bias or presupposition in the direction of a certain theory of value—one that I think many of us rightly regard as off-base and ultimately shallow. It is to distort the normative phenomena—to misdescribe the way the world presents to a valuer. The *world* and *other people* are generally what present to us as mattering, not our own states of mind.
- g. Returning to the problem we’re seeking to identify, and giving it rough expression from the first-person standpoint, the problem is that one day I will (or anyway might) lose everything. Suffering is also a problem, no doubt, and in many cases it would be good to get rid of it, but it’s not my *main* problem. I suffer *because* I’m aware that I’ll one day lose everything (or anyway might). We should get the phenomenon of valuing right from the get-go, else it might lead to other distortions as we build up from these foundations.¹⁴

¹² [Insert note about complications associated with translation. These are important points, but they don’t affect the point I’m making here.]

¹³ In “Constructivism in Ethics and the Problem of Attachment and Loss,” I discuss Kisa Gotami, a bereaved mother whose story is famous in the Buddhist tradition. She loses her child, goes nearly mad with grief, and receives help from the Buddha.

¹⁴ [Add note acknowledging the richness and complexity of the Buddhist tradition and teaching. There are of course many, many moments in which Buddhist descriptions of the problems we face emphasize the *losses* in question, as opposed to the suffering that we experience as a result of those losses. E.g. MN 82: Ratthapāla Sutta; II 65–82: “Master Ratthapāla, there are four kinds of loss. Because they have undergone these four kinds of loss, some people here shave off their hair and beard, put on the ochre robe, and go forth from the household life into homelessness. What are the four? They are loss through aging, loss through sickness, loss of wealth, and loss of relatives.” Nevertheless, I think there’s an unmistakable tendency that

9. Conclusion

- a. There is much left to argue about and clarify here. But suppose for the sake of argument you grant me that I've succeeded in identifying a universal, or at least very deep and widespread, problem associated with agency. Let me end by addressing, very briefly, some worries one might have in thinking ahead to "Step 3" of the larger constructivist project—which is to try to argue that the best solution to the problem in question involves seeking out and occupying a certain standpoint, which is, in turn, an ethical standpoint (or, more accurately, while not itself an ethical standpoint, a standpoint that transforms our agential standpoint *into* an ethical standpoint, assuming we "jump" back into agency).
- b. Sure, maybe you're right, you might say, maybe there's a nice solution there—good luck with that; it would be nice if so. But then you might rightly wonder: Aren't there, quite obviously, all kinds of *other* solutions to the problem of vulnerability to devastating loss, such that you'll never successfully argue that an ethical standpoint is the best or only solution? Some alternative solutions that immediately come to mind are: (1) denying the facts (thinking oneself invulnerable, believing one will be reunited with loved ones after death, telling oneself that everything, including the worst atrocities, happens for a good reason); (2) ignoring the facts by distracting oneself with work, pleasure, and entertainment; (3) hardening oneself (developing a shell, becoming cynical, cultivating invulnerability by shutting down emotionally); and (4) numbing out in various ways (treating the pain rather than the underlying problem—a familiar solution when the underlying problem is insoluble).
- c. This of course raises huge questions. I just want to close with the following, tentative suggestion. Yes, all those strategies *are* solutions. Moreover, I think each has its place on particular occasions. But as *general* policies—as one's default, go-to strategy for coping when there's a unbridgeable gap between "is" and "ought" from one's own point of view—each is sorely lacking on terms that are arguably involved in *any* (or at least many) agential perspectives. As general strategies, they are in tension with needs that one has from one's own point of view as a creature who is finite, but who cares about how reality will unfold, and who doesn't want to *give up* caring and trying to shape reality, even while knowing quite well that the odds against one's deepest hopes being realized are very high.

is at least accidentally associated with the Buddhist tradition to suggest that the fundamental problem is the experiential state itself, as opposed to the losses, or prospect of losses, that give rise to it.]

[Note: I'm planning to leave out of the talk the text with a line in the left margin (as well as the footnotes, of course).]

Official Corruption¹⁵

Niko Kolodny

In its broadest use, “corruption” means regress from a pure, healthy, or virtuous state. However, it has a narrower use, in which it means, roughly, “abusing an office for personal ends.” The paradigms of such “official corruption” include bribery, nepotism, cronyism, self-dealing, and embezzlement. My questions are what official corruption is, and why it’s wrong.¹⁶

My aim is to understand better which values justify our confident judgments about clear-cut cases. And I want then to apply that better understanding to resolve our ambivalence about hazier cases. Is it corrupt, for example, for a legislator to table a bill to attract campaign donations? Is it corrupt for a head of state to warm to a foreign power that projects his likeness on a hotel façade? The social science literature on corruption, extensive though it is, offers less help with these questions than we might have hoped. A recent primer, for example, just kicks the question back to us, proposing as a criterion of being corrupt: being “perceived by a significant proportion of the population and/or the state as corrupt” (Holmes 2015 17).

As with many moral-philosophical inquiries, the inquiring moral philosopher has an agenda. My agenda is somewhat parochially American. First, as American politics becomes ever more polarized, the condemnation of corruption remains a rare point of consensus. The word “corrupt” is less like “liberal” or “conservative,” and more like “undemocratic” or “unpatriotic.” One doesn’t need to know the speaker’s party affiliation to know whether she means it as a term of abuse. Our shared condemnation of corruption thus bespeaks shared values that, at least in principle, are levers for persuading the other side.

Second, there is Trump. The brazen entanglement of his Presidency with his vanity, finances, children, licensing business, and circle of misfit courtiers would have been, just a few years ago, scarcely imaginable in American politics. At the same time, “bribery” and “treason” are the only two “high crimes and misdemeanors” that the Constitution cites by name as grounds for impeaching the President. While it’s doubtful that Trump has committed treason—has waged literal war on the United States—it’s a far

¹⁵ [Thanks to UNC, especially Sophia Moreau, Humboldt, Northwestern, especially Paul Weithman.]

¹⁶ While the primary wrongdoer is the corrupt official, others, such as bribers, can also commit related wrongs. For example, offering bribes may abet the official in acting corruptly, or gain unfair advantage over others.

more open question whether he has been bribed. And it's not a question, as we will see, that the law settles on its own.

Third, money floods American politics. Some effort to stem the tide might seem in order. Since the 1970s, however, the U.S. Supreme Court has held that to regulate political money is to regulate constitutionally protected political speech. And the only valid rationale for that is combatting corruption or its appearance. Finding no substantial connection to corruption, the Court has thus struck down restrictions on campaign *expenditures* as opposed to contributions (*Buckley v. Valeo*), on contributions to *referenda* as opposed to candidate elections (*First National Bank of Boston v. Bellotti*), on contributions to groups that engage in expenditures *uncoordinated* with campaigns (*SpeechNow.org v. FEC*), and on *total contributions* to different campaigns by a single donor (*McCutcheon v. FEC*).¹⁷ In sum, the debate over campaign finance reform has become largely a debate over the meaning of “corruption.” Advocates of reform expand its meaning, at times almost into a catch-all for any shortfall from the ideal of good government (Teachout 2014, Lessig 2015 5–6, 247–50). Skeptics of reform, by contrast, insist on a narrow construal, no doubt gratified by the suggestion in *Citizens United* that corruption can only be *quid pro quo* and cannot consist in the purchase of access alone. Granted, moral philosophy would have little to add if this were a debate solely about how to apply settled law. But the debate seems, in part, free reflection on what corruption is.

My final reason for interest in corruption is less parochial, although more idiosyncratic. My current project is to explore the conjecture (or *idée fixe*?) that many of the ills of social and political life are the ills they are because of the relations of inequality that they perform or sustain. Anxieties about the maldistribution of economic gain (Anderson 1997, Scheffler 2003), about undemocratic forms of government (Kolodny 2014b), about discrimination on the basis of race or gender (Moreau, ms.), about illiberal constraints on personal choices (Kolodny ms. a), indeed about the very idea of the state itself (Kolodny 2016, 2017) are largely anxieties, I conjecture, not about resources, or opportunity, or freedom, or self-government, but instead about standing to others as an inferior. The same is true, or so I suggest, of our anxiety about corruption.

Sections 1–6 review the ideas that come soonest to mind: that corruption is wrong in some general way, such as fraud or theft; that corruption disserves the public interest; that corruption breaks the law; that corruption reflects a deficiency of civic spirit; that corruption violates the public's property rights; and that corruption plays favorites. After summarizing the limitations of these ideas, even when taken together, I make, in Section 7, a positive proposal, which I give the label, “Subordination by Personalization.” An official acts corruptly when he fails to exclude personalizing influences on his use of his office. And corrupt acts of that kind are wrong because they turn others' exposure to that office into a relation of inferiority to the natural person who occupies the office. Section 8 asks what does and doesn't count as corruption, according to Subordination by Personalization. Section 9 replies to criticism that the objection to corruption cannot be at root, as Subordination by Personalization suggests, an objection to inequality. Section

¹⁷ Indeed, in consistency, the Court should go further and lift the ban on the activities just listed, which it judges to be non-corrupting, by *foreign* entities (*Bluman v. FEC*).

10 distinguishes corruption, so understood, from several other objections to money in politics. However, section 11 concludes by noting that many of these other objections are, like the objection to corruption, rooted in an ideal of a society of equals. If so, then there is something incoherent in the Court's exclusive focus on combatting corruption, to the neglect of, say, equalizing access to the means of political influence. And it is further corroboration of the conjecture that many of our anxieties about political relations are anxieties about relating to others as inferiors, or, for that matter, as superiors.

1. Non-official wrongs?

Some corrupt acts involve "non-official wrongs": that is, wrongs that don't depend on their being official acts. Embezzlement, for example, is wrong, at least in part, because it is theft. It would be just as wrong, at least as theft, if a hacker drained the same sum from the treasury.

But not all corruption involves non-official wrongs. It isn't wrong, because it simply isn't possible, for someone who isn't a government official to award a government contract for a kickback. And even when corrupt acts do involve a non-official wrong, they seem wrong in a distinctive, further way, which depends on their being official acts.

This seems to be part of what is conveyed by the familiar slogans that corruption is wrong because it is "an unjustified use of office," or "subverts the public to the private," or "breaches the public trust." Suggestive slogans, to be sure. But what are they suggesting?

2. Public interest?

A natural answer is that officials have a duty to serve the public interest (perhaps simply because they have special opportunity to). By acting corruptly, an official disserves the public interest. (Friedrich 1972, Gardiner 1993, Philp 2002, Holmes 2015). For, in acting corruptly, the official acts for personal ends, and personal ends often diverge from the public interest. So, in acting corruptly, the official breaches that duty. That is how corruption "subverts the public to the private."

2.1. Definitions: "public interest," "office," "functions well," "sufficient sensitivity"

The "public interest," as I will understand it, consists in improving people's opportunities, making trade-offs among different people in a distributively fair way (e.g., prioritizing the worst off). Note that when an official action disserves the public interest, so understood, it is not some collective entity, "the public," that has a complaint. Instead, some member of the public has a complaint: namely, that a different official action would have improved her opportunities, in a way that would not have been unfair to anyone else. And note that what is to be distributed fairly is, in general, not the "consumption" of goods, but instead the opportunity to consume them: that is, a person's prospect of consuming them if she chooses appropriately. In general, the most that we can do is lead others to water; it's up to them whether they drink.

An “office,” as I understand it, is part of some larger system of institutions. It consists in an official making consequential decisions by certain processes. An office “functions well” insofar as it contributes, against the background of how the rest of the system operates, to the system’s serving whatever values the system is supposed to serve. These values include at least the public interest (as well as “equal influence” and “proper representation,” I argue in sections 10.1 and 10.2).

If the office functions well, it will be in part because the official follows decision-making processes that are “sufficiently sensitive” to certain “justifying factors.” This may be close to a tautology, since a factor “justifies” a decision insofar as the presence of that factor makes it the case that the decision, in virtue of its content, will contribute to the system’s serving whatever values it is supposed to serve. The fact that a contractor’s bid is low, for example, is a justifying factor for awarding him the contract. Something less than perfect sensitivity (a correlation of 1, as it were) can be sufficient, if greater sensitivity would be too slow, costly, unpredictable, or lacking in transparency.¹⁸

2.2. An illustration: nepotism

For illustration, consider the ways in which nepotism can disserve the public interest. An official, Offe, gives a job to his niece, Niecy. If another applicant, Abel, is better qualified, then John Q. Public, who relies on the job’s being done well, can complain that his opportunities could have been improved, by giving Abel rather than Niecy the job, without unfairness (since Abel’s claim is at least as strong as Niecy’s).

Furthermore, Abel can object that his opportunity could have been improved, by his getting the job, without unfairness to anyone. And even if Abel was only equally qualified, he may still object that the selection process, “Choose your niece,” was not sufficiently sensitive. A more sensitive process would have improved Abel’s chances of the job, without unfairness to anyone else. And even if Offe uses a sufficiently sensitive process, which leaves Niecy and Abel tied, Abel can complain if Offe then breaks the tie in favor of Niecy because she is his niece. Offe could have given Abel a better chance of getting the job, without unfairness to Niecy or anyone else, by drawing lots to choose between Abel and Niecy.¹⁹

Finally, if nepotism in this case leads people to expect nepotism in other cases, it may incentivize behavior that in turn disserves the public interest. Abel may divert his investments away from augmenting his own human capital into marrying into officials’

¹⁸ Very roughly, a process is “more sensitive” to justifying factors insofar as there is a higher correlation between the presence of justifying factors and the decisions that they justify. This is too rough, but it does bring out the point that the official need not be *aware* of justifying factors in order to be sensitive to them. He might be aware only of correlated proxies.

¹⁹ Scanlon (forthcoming) suggests that a lottery would be called for in this case not because (1) Abel has a claim to the better opportunity for the job that a lottery would give him. Instead, it’s called for because (2) it gives Offe a way to avoid “favoritism”: a “way of making a choice that does not involve giving” lesser weight to Abel’s interest. But (2) presupposes (1). *Which* interest of Abel’s does Offe give lesser weight to when Offe skips the lottery if not Abel’s interest in a better opportunity for the job?

families. And John Q. may hedge against officials' relatives' poor job performance by seeking substitutes, which itself may disserve the public interest.

2.3. *Why disserving the public interest is not the same as corruption*

To be sure, much corruption is wrong because it disserves the public interest. Yet if institutions are poorly designed (if, say, import duties are chokingly high), then the corruption of an official—call him “Bent”—might serve the public interest. Indeed, it's a serious, if contested, thesis that corruption can, under certain conditions, be economically efficient (Leff 1964, Nye 1967, Huntington 1968, Friedrich 1972). The rough idea is that corruption, by allocating resources to those most willing to pay, puts them to their most productive use.

In reply to this objection, that some corruption serves the public interest, one might put forward the following dilemma. If, on the one hand, Bent's switching to a corrupt decision-making *process* would serve the public interest better *in general or over the long run*, then there's no objection. Bent is just a reformer taking the initiative.

If, on the other hand, corruption disserves the public interest in general or over the long run, then it's just a lucky accident that Bent's corruption served the public interest *in this particular case*. Bent acted wrongly because he defied a duty to follow a sufficiently sensitive decision-making *process*. Why did Bent have such a duty? Perhaps because Bent has a duty to take due care to get things right. Or perhaps because others were entitled to opportunities that *partly consist* in Bent following sufficiently sensitive processes. (When we say, for instance, that Abel was entitled to a “fair shot” at the job, that just means that Abel was entitled to Offe's following a sufficiently sensitive process in awarding it.)

Against the first horn of the dilemma, however, there's still an objection to Bent's unilaterally adopting a long-run policy of taking bribes, even if it's more efficient. Against the second horn, suppose Bent follows a sufficiently sensitive process, but it underdetermines the decision in this particular case. There's still an objection to Bent's taking a bribe to break the tie in the briber's favor in this particular case.

Moreover, even if all corrupt actions do disserve the public interest, not all actions that disserve the public interest are corrupt. Even non-corrupt officials can louse things up. But lousing things up isn't wrong, at least not in the same way (Philp 2002).

3. Law?

Even if corruption doesn't disserve the public interest, it may still break the law (Friedrich 1972, Gardiner 1993, Nye 1967, Leff 1964). Breaking the law may be wrong, for any number of reasons. In particular, if Bent's law-breaking is wrong because he led others to expect (say by taking an oath) that he would follow the law, then it literally “breaches the public trust,” as corruption is often said to do (Lowenstein 1985 806, Philp 2002).

Not all corruption breaks the law, however. Many corrupt acts are within the official's legal discretion. The law places few constraints on what laws legislators can

make, for example. Indeed, the corrupt act may be legally required. Bent may be bribed to do his job.

To be sure, corruption may break laws that specifically prohibit corruption itself. But corruption would be wrong even if there were no law against it. Arguably, that's why there *are* laws against it. Moreover, bribery law in the U.S. appeals to extra-legal, moral standards to determine what counts as bribery (Lowenstein 1985, 2004).

Finally, just as garden-variety poor performance can disserve the public interest, so too can it break the law. But again it's not wrong in the same way as corruption.

4. Motive?

This recurring contrast between corruption and mere poor performance might suggest that the wrong lies not in Bent's "outer" action, but instead in his "inner" motive (Ryan 2013, Teachout 2014). The private subverts the public, on this view, when self-interest contaminates civic spirit.

But which motive makes for corruption? Believing one is dis-serving the public interest (Raz 1977 on "arbitrary power"; Teachout 2014 289) is not necessary for corruption. When Bent breaks the tie, he knows that he isn't dis-serving the public interest.

Nor is acting from private interest sufficient for corruption. When don't politicians act from private interest? Consider how Zephyr Teachout's proposal to use the criminal law to prevent corruption risks backfiring, if corruption, as Teachout at times suggests, consists in acting from private interest (2014 285). For if politicians act from fear of punishment, they are already acting from private interest. Far from preventing corruption, the criminal law dangles an almost irresistible temptation to it.

Moreover, whatever Bent's motive, why does he wrong anyone by acting from it? While I care that "Representative Barbara Lee speaks for me," why should I care, so long as she does "speak for me," what hidden springs set her tongue in motion? After all, it hardly stokes resentment to learn, from Adam Smith, that it's not from the benevolence of butcher, brewer, or baker that we expect our dinner.

And who is wronged by the corrupt act? The people who are somehow "targeted" by the corrupt motive? But our nepotist, Offe, may give no more thought to Abel than does his non-corrupt, but absent-minded counterpart.

5. Property?

Some corruption is wrong because it violates the public's property rights. Imagine Bent "rents" out the township's snowplow and pockets the proceeds. Here the "public is subverted by the private" insofar as the public's property is used for private gain.

Granted, not all corrupt acts involve the use of something naturally described as "property," such as materials, equipment, physical plant, patents, copyrights, emissions

permits, broadcast frequencies, and so forth. Still, one might suggest that the office itself is the public's property, and so any use of the office must respect the public's property rights in it. Among those property rights, it might be said, are rights to any "proceeds" from its exercise.²⁰ Recall the suggestion in section 2.3 that where Bent's bribe-taking allocates resources more efficiently, we should commend him for forward-looking institutional reform. One might now reply that even if the allocation is more efficient, the resulting "social surplus" is not Bent's to take. Even if Bent's act is not otherwise wrong, it still violates the public's property rights for him to be "enriched" by its exercise (Strauss 1995).

However, some corruption, such as Offe's nepotism, doesn't enrich (apart from "avuncular gratification," I suppose). And some enrichment from office isn't corrupt. Suppose that someone enriches Bent, in a way that Bent cannot disgorge, in the hope of influencing him, but Bent refuses to be influenced.²¹

6. Favoritism?

6.1. The anti-favoritism principle

In recent work, Scanlon (forthcoming) explores an idea that, while not billed as an explanation of the wrong of corruption, might be put to that purpose. To a first approximation, his principle of "equal concern" says that, for certain kinds of relationships, if an agent, A, gives a benefit, B, to X, and A stands in a relationship of one of those kinds to X, and A *also* stands in a relationship of the same kind to Y, then A owes it to Y to give B to Y. To distinguish equal concern from other egalitarian principles, I take the liberty of calling it "anti-favoritism." It is, one might say, the philosophical distillate of the kindergarten idea that one shouldn't share a treat with some classmates unless one "brought enough for everyone." Paradigms of the relationships that call for anti-favoritism are those of parent to children and of state to citizens. While we object when a state fails to provide to some citizens the public services that it provides to others, Scanlon observes, we take it for granted that a state will "fail" to provide to non-resident foreigners the public services that it provides to its own citizens.

It is crucial to distinguish Y's complaint about A's favoritism toward X from Y's complaint that A didn't benefit Y when A could have. Y's favoritism complaint is "comparative": it depends on what A *actually does for X*. Y's "didn't-benefit" complaint, by contrast, is "non-comparative": it depends on what A *could have done for Y*, given the physical and moral constraints A faced (including the moral constraint that doing more for Y would have come at X's expense in a distributively unfair way). We will soon consider several other ways in which these complaints can come apart. But for

²⁰ Among these rights, it might also be said, are rights to decide what is to be done with the office. But this seems to boil down to the suggestion in section 3 that what's wrong with corruption is that it violates the law.

²¹ A more difficult case to evaluate, to my mind, is a retired official's voluntary acceptance of royalties from the sale of her memoirs of her time in office (compare Thompson 1995, ch. 3). Even if one objects to her using her office to enrich herself, one can hardly object to her making a living from a second career as a writer.

now, suppose that A has a noncomparative duty to give B to Y, but A has not actually given B to X. Then A may still have no comparative duty to give B to Y.

This is only a first approximation, as I say. Anti-favoritism can't strictly require that whenever A gives B to X, A must also give B to Y. For X and Y might not be similarly situated. For example, X might need B more. Scanlon suggests that giving B to X but not to Y does not violate anti-favoritism when doing so would be justified "if the interests of all those affected were given appropriate weight." But this leaves anti-favoritism all but toothless. When does A give X's interests "appropriate weight"? When A complies with whatever *requirements* X's interests justify? Suppose that X's interests neither require nor prohibit giving B to X and Y's interests neither require nor prohibit giving B to Y. Then, on the present suggestion, giving B to X but not Y would satisfy anti-favoritism, since it gives X's interests and Y's appropriate weight.

Properly qualified, I think, anti-favoritism should instead read: for certain kinds of relationships, if an agent, A, gives a benefit, B, to X, and A stands in a relationship of one of those kinds to X, and A also stands in a relationship of the same kind to Y, then A owes it to Y to give B to Y, *unless some difference between X and Y justifies not giving B to Y*. When giving B is an official act, a justifying difference will presumably be a difference in justifying factors. The low cost of a contractor's bid is a justifying factor for awarding a contract, but the contractor's old-world charm or promise of a kickback is not. As these would not be justifying factors even if no other contractor were to bid, they don't make for justifying differences if another does.

6.2. *The difference comparative complaints make*

Enumerating the ways in which Offe's nepotism disserves the public interest, we listed, *avant la lettre*, noncomparative complaints that Abel might have: complaints that would apply even if Abel were the only qualified applicant and Offe let the position go unfilled. Anti-favoritism, however, gives Abel a further, comparative complaint when the job goes nepotistically to Niecy. This comparative complaint can make a difference even in such cases, where a noncomparative complaint already applies. For one thing, the comparative complaint is often easier to establish. It is hard to show that a given regime of road maintenance, for example, best serves the overall distribution of opportunity. Perhaps everyone would be better off with pedestrian malls and light rail, instead of roads. By contrast, it's easier to show that the potholes in X's neighborhood but not in Y's were filled and that, save for their racial composition, the two neighborhoods are similar.

Moreover, comparative complaints sometimes apply when noncomparative complaints do not. First, even if A could improve Y's opportunity by giving B to Y, A might not have a duty, or a duty to Y, to do so. Although A would not noncomparatively wrong Y by not giving B to Y, A might nevertheless comparatively wrong Y by giving B to X but not Y.

Second, Abel might complain if Offe nepotistically gives the job to Niecy, without so much as considering Abel, even when Offe would have given the job to Niecy

anyway, since she is in fact better qualified than Abel. Here, Abel has no noncomparative claim on Offe to the *better chance at the job* that he would have gotten if Offe had considered him, since he would not have had a better chance even if Offe had considered him. Nor, presumably, does Abel have a noncomparative claim on Offe to *consider him for its own sake*. It is not as if there is standing reason for Offe to create openings so that people can have their qualifications reviewed. However, Abel has a comparative claim on Offe to consider him, given that he considered Niecy. Offe's nepotistic neglect is objectionably unequal *treatment*, even if it is not an objectionably unequal *distribution of opportunities*. As we will see in section 6.3, the justification of anti-favoritism implies that it should forbid such unequal, if otherwise inconsequential, treatment.

Finally, suppose that A gave only B to Y in the past, but now faces a choice whether to give more than B to X. In particular, imagine that while A applied a certain procedure in Y's case in the past, A faces the question of whether to apply a different procedure in X's case now, "bending the rules" in X's favor. Y has no noncomparative complaint about what A now does for X, since that has no bearing on what A could have done for Y. Still A might feel that it would be somehow unfair, presumably to Y, to bend the rules now for X. (At any rate, I often feel this way, when, as a teacher or administrator, I'm asked for an extension, waiver, exception, etc. that I've denied before.) Anti-favoritism would explain why. Anti-favoritism might thus answer some versions of the charge of "rule worship": Why adhere to a rule, when violating it in a particular case does no harm to the purposes that the rule is supposed to serve?

6.3. When and why does the principle of anti-favoritism apply?

Which relationships prohibit favoritism? Put another way, what do the parent-child and state-citizen relationships have relevantly in common? Scanlon suggests that these are relationships in which one relative is obligated (independently of anti-favoritism) to provide the other with benefits.²² But Scanlon doesn't explain why being obligated to benefit should trigger a further requirement of anti-favoritism. And this suggestion would make anti-favoritism apply to both too few and too many cases, by Scanlon's own lights. Too few: Scanlon holds that anti-favoritism applies to the state's provision of recreational facilities, while not insisting, as far as I can tell, that the state is *obligated* to provide recreational facilities.²³ Too many: Scanlon holds that anti-favoritism does not apply to your individual contributions to global humanitarian aid. If you send additional

²² The requirement of equal concern "presupposes an obligation on the part of some agent to provide benefits to everyone in a certain group" and "applies to agents who have an obligation to provide benefits to certain people." At times, Scanlon goes so far as to suggest that equal concern applies only "to inequalities that result from the failure of that agent to fulfill [an independent] obligation to all those to whom it is owed." But this can't be what he means. For in his main examples, while A has noncomparative obligations to give B to X and Y, and while A does not fail to fulfill *these* obligations, A then violates equal concern by giving X, but not Y, some further increment of benefit.

²³ Another worry is that it violates equal concern for the state to punish or conscript some, while letting others off the hook, although neither punishment nor conscription involves any obligation to *benefit* convicts or selective-service registrees.

aid to group A, you are not required to top up your aid to group B. But I doubt that Scanlon wants to deny out of hand that your contributions discharge *obligations* to aid.²⁴

So I suggest a different answer. A's treatment of X and Y is subject to anti-favoritism just when and because, first, equal standing between X and Y matters and, second, A's favoritism of X over Y would threaten (whether causally or constitutively) this equal standing. It is important, for instance, that a child be secure in the knowledge that it comes after no other child in the affections of its parent. And it is important that, whatever hierarchy there may be in other social relationships, a citizen stands as an equal, as a citizen, to other citizens.²⁵ When the state favors some citizens over others, it confirms the latter as "second-class" citizens. Think, for example, of FEMA prioritizing aid to Texans over to Puerto Ricans.

Thus, anti-favoritism applies to any form of unequal treatment that bears on equal standing. This treatment might be something other than the distribution of opportunities, such as when Offe nepotistically neglects Abel's application, even when Abel would have lost out to Niecy anyway. When others are considered, it matters that you be considered too.

6.4. *Corruption and favoritism distinguished*

We will say more about this in section 10.3, in connection with the problem of money in politics. For now, though, what's important is that corruption and favoritism are distinct. Granted, corruption often involves favoritism. When Offe does something for Niecy, or Bent does something for Briber, there is often indeed someone else on the scene for whom they do not do the same, even though there is no justifying difference between them.

However, corruption doesn't always involve favoritism. Bent might withhold a benefit from X until paid a bribe, even though there is no other similarly situated Y on the scene. Or A might demand and receive a bribe from everyone.²⁶

²⁴ Still more puzzlingly, Scanlon suggests that if some global institution held these obligations of humanitarian aid, then equal concern would apply to that institution. Whatever other differences there may be between the individual and the global institution, however, both may have *obligations* to give humanitarian aid.

²⁵ Scanlon does offer an argument for why anti-favoritism applies in the special case of state and citizens. "[W]hy should some citizens accept that the interests of others count for more than their interests in justifying government policies, which they are being asked to cooperate in supporting through taxes and compliance with other laws?" However, this argument says nothing about the state's obligations to benefit its citizens. And is the argument that if the state provides a benefit to X, but not to Y, then it somehow doesn't compensate or reciprocate Y's burdens of taxation or compliance? But this argument has little to do with equal benefits. If the state used taxes for some good that failed equally to benefit X and Y, the state would similarly fail to compensate or reciprocate their burdens of taxation.

²⁶ One might reply that it is enough for a complaint of favoritism that A *would* not give B to some similarly situated Y *were* they to appear on the scene. But is there a complaint? Who has it? Not the similarly situated Y, since there is no similarly situated Y.

And favoritism doesn't always involve corruption. FEMA can favor Texans simply because it is mismanaged.

7. Subordination by personalization

7.1. Subordination by personalization

So, what is corruption, and why is it wrong? My positive answer, such as it is, starts from the observation that to be subject to an official's decision is to be subject to the greater power and authority of the office. By "authority," I mean "de facto authority": simply the ability to issue commands that are, for whatever reason, generally complied with. (Exercises of de facto authority may not give anyone reasons, let alone obligations, to comply.) In general, to be subject to the greater power or authority of another individual presents a moral problem. A variety of moderating conditions, to be sure, can address the problem. But absent those moderating conditions, to be subject to the greater power or authority of another individual is to stand to him as an inferior to a superior, to be subordinated to him (Kolodny 2014b). And one has an objection to being so related to another, or so at least a resonant strain of egalitarian thinking suggests.

One such moderating condition, I think, is the key to corruption. It is that when the official decides how to use the office, and so its superior power and authority over others, he excludes personalizing influences. A personalizing influence, P, on the decision of a natural person, S, occupying an office, O, is an influence (i) that expresses the practical or emotional attitudes of the particular, natural person who S is, (ii) where influences like P are not necessary for O to function well (i.e., not necessary for an occupant of O to be sufficiently sensitive to justifying factors), and (iii) where P "makes a difference": that is, where an occupant of O influenced by P would decide at least some cases differently from an occupant of O who excluded all influences satisfying (i) and (ii).

An influence is "excluded" when it is kept it from affecting the decision-making process. (An influence can "affect" a decision-making *process* even if doesn't *change* the *decision*: even if the same decision would have been made even without that influence.) Such exclusion can take place in several ways. Factors that would have triggered the influence may simply not arise (which can be engineered, e.g., by an official's being recused from cases that would present a conflict of interest). Or the decider may be kept ignorant of those factors (e.g., by screens, blind trusts, redactions, sequestrations, conferences in chamber). Or, as a last line of defense, the decider may resist the influence of such factors on the decision-making process (e.g., by giving such factors no weight as reasons, as Raz's "exclusionary reasons" require). In various contexts, for various reasons, deciders may be expected to exclude influences that are irrelevant to the decision, or that would discriminate against race or gender, or that some authority has instructed them to disregard. The specific suggestion here is that officials, in general, should exclude personalizing influences, and that they should do so in order to avoid or moderate the relations of inferiority that the superior power and authority of their offices would otherwise (causally or constitutively) threaten.

But why should excluding personalizing influences avoid or moderate such relations of inferiority? The idea is that when an official, call him “Ness,” excludes such personalizing influences, he leaves no impression on the decision, or at least a fainter impression, of the natural person he is. Either the particularities of his practical and emotional attitudes don’t affect his decision-making process, or anyone in that office (if it is to function well) will be influenced by particularities of that kind, or those particularities make no difference. Consequently, one wants to say, it might as well be anyone wielding the power and authority of the office.²⁷ To be sure, a natural person, Ness, decides. But that natural person decides by and through an impersonal persona, on which, as it were, features and expressions have been fixed.

Thus, the exercise of the power and authority of the office is not (or at least less) the act of a particular, natural person, and instead the exercise of the power and authority of something that is no other, particular, natural person—the office, one wants to say, or the law, or the state, or the general will. This avoids or moderates the threat of inferiority or subordination, because inferiority or subordination is possible only to another natural person. We have no similar claim to equal standing with things that aren’t natural persons, such as offices, institutions, laws, or collectivities. The point is not that these entities have superior standing, but instead that assertions or denials of superior, inferior, or equal standing with these entities make little sense. What would it be, for instance, for me to be the equal of Berkeley, California, the United States of America, and the United Nations? (If equality is transitive, would Berkeley then be the equal of the USA?)

This is one way of spelling out the importance of the distinction between the impersonal role and the person who occupies it, a distinction which, as has long been noted, is central to the “modern Western” conception of the state (Weber 1956 ch. 11, van Klaveren 1957, Huntington 1968). And it is a way of spelling out what we (at least sometimes) mean by, and value in, the rule of law, especially when we contrast it with the rule of men.²⁸

When, by contrast, the occupant of an office uses its special power and authority (whether by exercising it or otherwise²⁹) without excluding personalizing influences, the threat reasserts itself: that those subject to office are subordinated to the natural person who occupies it. That use of office wrongs them, for that reason. According to this account, which I give the cumbersome title, “Subordination by Personalization,” corruption consists in just such uses of office, uses which include at least the paradigm cases of corruption: bribery, nepotism, self-dealing, cronyism, embezzlement. The sense

²⁷ To say that the official “might as well be anyone” in this sense isn’t to say that discharging the office is something that “just anyone” could do. Few of us may be equal to the rigors of the office—not least the strain of deciding in a way that cordons off the natural person who he is. And we might rightly praise him, the natural person, for discharging the office well.

²⁸ Contrast Raz (1977) who suggests that nothing of interest can be made of this phrase. This is not to deny that what is often meant by the “rule of law” is captured by the ideal that Raz describes, whose core is enabling the conformity necessary for law’s achieving its “direct” purposes.

²⁹ Accepting bribes without giving anything in return, for example, is a use, but not an exercise, of office.

in which corruption “subverts the public to the private” is that it turns subjection to the public office into subordination to the private person who occupies it (compare Philp 1997).

Subordination by Personalization explains why only official acts can be wrong in the distinctive way that corruption is. The demand to exclude personalizing influences, which corruption violates, applies only to decisions about how to use asymmetric power and authority, where other moderating conditions are absent. For it is only such decisions that present a threat of inferiority. For example, Abel has no objection, on the grounds of corruption or otherwise, that Offe should decide, precisely on the grounds that Niecy is his niece, to send her, but not Abel, a birthday card.

Furthermore, even within an asymmetry of power and authority, the demand for exclusion is less pressing when other moderating conditions are already present. First, the asymmetry may exist only momentarily, as part of a one-off encounter, instead of an ongoing relationship within an established social structure. Second, the asymmetry may be limited to certain times, places, or contexts. Third, the power and authority may be limited in content: that is, in what can be done or commanded. Fourth, the asymmetry may be escapable, with little cost or difficulty. Lastly, the asymmetry may not be “final”; a higher court of appeal may govern how the power and authority are exercised.

Such other moderating conditions are typically absent, however, within asymmetries between officers of the state and those subject to their offices. The state cannot be escaped, for one thing, and it is the highest court of appeal (short of heaven). So the demand for exclusion tends to be more pressing for officers of the state, especially as we ascend the hierarchy of offices. What matters, however, is not whether the office is one that we are in the habit of thinking of as part of the “state,” but instead whether other moderating conditions are absent. The demand for exclusion may be just as stringent in certain “privatized” offices, such as that of warden in a for-profit prison.

Subordination by Personalization explains, further, why neither disserving the public interest, nor breaking the law, nor enriching, or nor favoring is necessary for corruption. It’s sufficient that personalizing influences are not excluded. And there can be personalizing influences on actions that don’t disserve the public interest, break the law, enrich, or favor.

Subordination by Personalization also explains why neither disserving the public interest, nor breaking the law, nor enriching, or nor favoring is sufficient for corruption. Personalizing influence is necessary. And actions can disserve the public interest, break the law, enrich, or favor even if personalizing influences are excluded. We can thus distinguish corruption from standard-issue poor performance.

7.2. Motive revisited

According to Subordination by Personalization, an official acts corruptly only when a personalizing influence affects her decision (for it is only then that she has failed to exclude personalizing influences). In some cases, moreover, the very same decision, had

it not been affected by a personalizing influence, would not have been wrong. (Take Offe’s giving the job to Niecy because she is his niece, when she would have been hired on the merits.) But then isn’t this just to say that the official’s motive is what makes corruption wrong? And isn’t this just the appeal to a bad motive—to self-interest or deficient civic spirit—that we supposedly found wanting in section 4?

To gauge our progress, recall that we were then at a loss to say what objection others have to the bad motive. Why should we have any more claim to the benevolence, say, of our officials than to that of our butchers, brewers, or bakers? By contrast, we do have a claim that officials not make the power and authority that they wield over us *their* power and authority as natural persons. And that is a claim on them to exclude personalizing influences in the use of their offices. We can also now say who, in particular, has an objection to Bent’s corruption. They need not be “targeted” by Bent’s motives. They are simply those over whom Bent’s office exercises power and authority—which might be, if Bent holds high enough office, every citizen.³⁰

We can also explain how criminal penalties for corruption can be something other than entrapment. Suppose Dieter routinely declines bribes simply because he fears the criminal penalties for bribery. This fear is not a personalizing influence, because, in the sense defined in section 7.1, it “makes no difference.” If a counterpart to Dieter, Ness, were not sensitive to *any* personalizing influences—that is, if he were uncontroversially non-corrupt—he would decide, in all cases, in the same way. Thus, even when Dieter decides out of fear, he leaves no impression of the natural person he is, because he would decide, in all cases, in the same way as someone, Ness, who, we agree, would leave no such impression.

To draw a final contrast with the appeal to bad motive, notice that all manner of bad motives may lurk behind the successful exclusion of personalizing influences. Although Dieter does not act corruptly, his motives hardly speak well of him. A better person, like Ness, wouldn’t need the threat of jail time to stay straight. We might take a similarly dim view of a non-corrupt official who would now be using his office to boost the value of his portfolio, if only he hadn’t, during the primary campaign, placed his assets in blind trust in order to remove the advantage of his opponent’s well-earned reputation for propriety.

8. What counts as corrupt?

8.1. Expanding what counts as corrupt

Taking personalizing influences to be sufficient for corruption tends to *expand* the class of corrupt acts. For one thing, the corrupt act need not be *quid pro quo*: an explicit, pre-arranged agreement that a specific official act will be performed in return for a specific personal favor.³¹ It may be instead repaying a debt of gratitude, or succumbing to flattery

³⁰ It follows that Niecy herself is a “victim” of Offe’s nepotism, at least if she did not consent to it. If this sounds odd, I suspect, it is because it is, in general, odd to describe someone as a victim of wrongdoing when they have reason to welcome the wrong done to them.

³¹ The applicable federal laws governing bribery reflect at least two different dimensions of “quid-pro-quo-ness”: how specific the official act and how definite the agreement to perform it. A conviction is

(Teachout 2014; Lessig 2015 91–99). Such influences also express the practical or emotional attitudes of the natural person occupying the office. Moreover, there can be personalizing influences that the official is not even aware of (Lessig 2015 98), let alone intends (Lowenstein 1985 795–6). Self-deception or obtuseness is no defense.³²

8.2. *Limiting what counts as corrupt: “weak states”*

Subordination by Personalization has a countervailing tendency, however, to *contract* the class of corrupt acts. This is because an influence is not a personalizing influence, by the definition of section 7.1, if influences *of that kind* are *necessary* for sufficient sensitivity to justifying factors: that is, necessary for the office to function well, against the background of the rest of the system.

This exception may explain why certain charges of corruption feel misplaced. Consider Fukuyama’s (2011 x–xiii) description of present-day Papua New Guinea. Each local “wantok” expects that its “Big Man,” once elected to the national parliament, will simply try to siphon off as much public spending for it as he can. Pork-barrel politics is all there is. Should we condemn this as “corrupt”?

We may be unsure, because we may be unsure whether there is a better alternative. First, if there is little trust among different wantoks, and if there are strong perceived obligations to favor one’s wantok, then each Big Man’s looking out for his own wantok, within the right sort of competitive structure, may be the decision-making process, of those realistically available, most sensitive to justifying factors. Exhorting Big Men to take a broader view may be counterproductive.

Second, among the opportunities that the system is supposed to provide might be opportunities for valuable relationships. When officials exclude personalizing influences, however, they can no longer have relationships, such as that between a Big Man and his wantok, that require such personalizing influences. These extinguished relationships may have been valuable in themselves. Or they may have been valuable for their effects. (Whether or not trust, loyalty, and esprit de corps are good in themselves, there’s no doubt that they grease the wheels.) The extinction of these relationships may be too high a price to pay for reforms that otherwise would better promote opportunities. By contrast, our own culture makes available personal relationships, such as forms of friendship, that don’t require officials to use their offices to favor their relatives. So the withering away of personal relationships that do require it is far less of a loss.

8.3. *Limiting what counts as corrupt: personal outlook*

less likely if the official only vaguely agrees to help the briber when she can, or agrees only to give some, perhaps not dispositive, weight to the favor in deciding whether to perform the act (Lowenstein 2004).

³² Arguably, even laziness may be corrupt. The personalizing influence in such a case would not favor a particular *decision*, in the way that nepotism favors hiring one’s niece. But it would disfavor certain kinds of *decision-making*: namely, kinds that require effort. That too may express the practical or emotional attitudes of the natural person.

For similar reasons, it may be misplaced to condemn certain officials for corruption simply because they are guided by what we might call their “personal outlook.” Take it as given that offices such as legislator or justice of a constitutional court function well only if those officials deliberate about matters for which there are no settled, agreed criteria, such as about what serves the public interest or about how the law is to be interpreted. And take it as given that such deliberation is *inevitably* guided by one’s personal outlook: one’s sense of where the public interest lies or how the law is to be interpreted. Then being guided by one’s personal outlook will not be corrupt, simply because it will be part of *any* decision-making process sufficiently sensitive to justifying factors.

8.4. *Limiting what counts as corrupt: strategic decisions*

More surprisingly, perhaps, many “strategic decisions” turn out not to be corrupt.³³ Consider offices that function well insofar as officers decide on the basis of judgments, reached by deliberation in certain conditions, about what will serve certain general aims, such as the public interest. They make “strategic decisions” when they make such a decision not on the grounds that it will immediately serve those general aims, but instead that it is a means to serve those general aims better overall or over the longer term.

Suppose, to fix ideas, that the office of legislator functions well when legislators decide on the basis of judgments, in the right deliberative conditions, about what will serve the public interest.³⁴ Legislator Lexi might judge that in order to secure passage of a bill that will serve the public interest, she must remove a provision, opposed by the last remaining holdout, that would have served the public interest even better. This strategic decision hardly seems corrupt.

But then how is it different for Lexi to remove the same provision in order to secure a donation necessary for her reelection? She may judge that that strategic decision likewise best serves the public interest.³⁵ After all, who better to serve the public interest as Lexi sees it than Lexi herself (compare Lowenstein 1985 845)? In a perfect world, perhaps, Lexi would, as Thompson recommends, only “seek contributions from citizens who want [her] to do what [she has] good reason to do” (1995 117). The trouble is that, in our imperfect world, modifying positions to attract contributions from citizens who don’t want Lexi to do what she has good reason to do may nonetheless help Lexi do what she has good reason to do.³⁶

³³ Thompson takes this condition, that the bribe is some purely political advantage, to distinguish “institutional corruption.”

³⁴ Alternatively, we might say: “best prospect of satisfying standards of proper representation.” In that case, the legislator would be a “delegate,” rather than a “trustee.”

³⁵ Lexi may also want to be reelected, simply as a matter of personal ambition. This would not make her use of office corrupt, so long as it does not make a difference: so long as it would not lead her to use her office in ways other than the public interest, as she sees it, recommends.

³⁶ For similar reasons, it’s harder to make the charge of corruption stick to partisan gerrymanders, which increase the party’s representation and so further its aims, than to bipartisan, incumbent-protection

8.5. *Explaining consensus*

We have observed several ways in which what counts as a personalizing influence, and so corruption, depends on what it is for that office to function well, against the background of the rest of the system, which in turn depends on the values that the overall system is supposed to serve. But since people so often disagree about those values, why should they so often agree about what counts as corrupt?

The answer is that often they don't need to agree on what it is for the office to function well, so long as they agree that whatever it is for the office to function well, it doesn't require the official's being influenced in the way at issue. For example, whether or not decisions to lease public land should be sensitive to environmental or business concerns, all can agree that they need not be sensitive to whether the lessee is the official's niece.

9. **Why inequality?**

Subordination by Personalization holds that the objection to corruption is, at root, an objection to inequality: to the asymmetric relationships between officers and those subject to their offices. There are, or so I have argued, good reasons for taking this view. Among other things, it is hard to explain why else the objection should apply, specifically, to uses of office. What makes offices special, if not the asymmetries they involve? Nevertheless, this focus on inequality might seem open to challenge, on several grounds.

9.1. *Can't inegalitarians condemn corruption?*

For one thing, one doesn't need to be an egalitarian to condemn corruption. However, Subordination by Personalization is broadly compatible with two inegalitarian stances. Consider first an inegalitarian who thinks that a given asymmetry of power and authority needs no justification whatsoever, beyond a reminder of the order of things. It would be compatible with, indeed would confirm, Subordination by Personalization that this inegalitarian would see no valid complaints of corruption within *that* asymmetry, while finding valid complaints of corruption in other asymmetries, that *do* call for justification. Thus, a Roman of senatorial rank might object to being compelled to do personal favors for some upstart, freedman bureaucrat, while not objecting (memories of the republic having dimmed) to being compelled to do personal favors for the Emperor, much less thinking that his slaves had any objection to being compelled to do personal favors for himself.³⁷

gerrymanders, in which incumbents sacrifice their own party's aims in order to draw safe seats for themselves.

³⁷ Indeed, some cross-cultural studies suggest that perceptions of corruption are sensitive, in this way, to perceived status relations:

In Sicily, as in Greece, it is regarded as legitimate for officials to require bribes or patron intervention in order to give the peasant what should be his due according to the law. In contrast to the modern setting, where even lower-class citizens are infused with some measure of egalitarian ideology, the peasants in this kind of community accept the arrogance of officials as an almost legitimate function of their higher social status and regard it as natural that their sympathy can only be aroused through gifts and bribes (Heidenheimer 1970).

The second inegalitarian stance, by contrast, accepts that asymmetries of power and authority are a problem. And it accepts that part of the solution is that officials perform their roles impersonally, excluding personalizing influences. However, it denies that some further, egalitarian structure, such as equal influence, must also be part of the solution. A non-corrupt bureaucracy suffices; democracy is not necessary. This stance accepts Subordination by Personalization. If it disagrees with anything this paper, it is with the speculation later on that the objection to corruption may be linked to objections to certain further, egalitarian structures, such as equal influence.

9.2. *Cheating*

A second challenge to Subordination by Personalization is that corruption appears to occur in symmetric relationships. In penetrating comments, Sophia Moreau imagines that a lawyer in a civil case, Saul, doesn't disclose documents to the opposing lawyer, Dooright. It's natural to describe Saul as acting corruptly, and it is natural to say that this corruption wrongs Dooright. However, Saul is not exercising any *superior* office to which Dooright is subject. Dooright is an *equal* adversary.

This is, however, simply a different kind of wrongdoing—namely, cheating—from that exemplified by paradigms of official corruption. Cheating presupposes a competitive process that has value (or whose outcome has value) only insofar as the relevant actors (including the competitors, referees, and so on) refrain from taking certain “illicit” means to bring it about that a given competitor wins (or achieves what, if not for such cheating, would count as winning). When some such actor takes illicit means, this deprives the competition (or its outcome) of value. So, it does, or can, wrong everyone who has an interest in that value.

On the one hand, some cheating isn't corruption, because it doesn't involve official acts. To be sure, in the case of civil suits, it is natural to describe the competing attorneys as officials (“officers of the court”). So, it is natural to describe Saul's cheating as an “abuse of office.” And so, given our initial, idea-fixing definition of official corruption as an “abuse of office,” it is natural to describe Saul's cheating as official corruption. But cheating also occurs in games and sports, where the competing players, at least, are not naturally described as “officials.” On the other hand, some corruption isn't cheating, because it doesn't spoil a competition. Bent's corruption did not wrong a competitor.

Furthermore, corruption depends on what influences the decision. For example, making a decision for a bribe is corrupt whereas making the same decision on the merits is not. By contrast, whether cheating took place does not depend on what influences the decision. So long as Saul (intentionally or negligently) takes illicit means, it doesn't matter whether he takes them from self-interest, incompetence, or sincere conviction (e.g., that privacy rights of his client shield her from such disclosure).

9.3. *Selling votes*

Another apparent case of corruption within a symmetric relationship is one citizen's selling his vote to another citizen, in a context where every citizen has the same votes to sell and the same money with which to buy them (compare Straus 1995). Suppose Sellars knows that two candidates, Candi and Candy, are tied for first, with respect to the public interest (and any other relevant values). Without disserving the public interest, he accepts payment from Kaufmann to vote for Candi over Candy. Yet neither Sellars nor Kaufmann has superior power or authority over the other.

If this exchange is wrong, it is again wrong for different reasons from those for which our paradigms of official corruption are wrong. On the one hand, if it isn't public knowledge that people may and will buy and sell votes (as it presumably isn't where the secret ballot was adopted precisely to discourage vote buying), then the back-alley deal between Sellars and Kaufmann is wrong simply because, as we discuss in section 10.1, it violates equal opportunity for influence over political decisions. Kaufmann is getting an extra vote by means that others, even with the same money to spend, assume is ruled out.

On the other hand, if it is public knowledge and common practice that people buy and sell votes, then our evaluation of what Kaufmann and Sellars are doing, I suspect, turns on our evaluation of the system in which they are participating. If we became convinced that such a system would serve the public interest (because, say, it tracked intensity of preference), then we would view it simply as an unfamiliar, but not objectionable, way of conducting elections.

9.4. *Isn't corruption just domination?*

Neo-Roman "republicans" might suggest that we can avoid the counterexamples considered in this section by understanding corruption not as subordination, but instead as "domination," which (despite the word's origin and standard meaning) does not require inequality. You are "dominated," insofar as an alien will can, even if you know that it will not, arbitrarily interfere in your choices (Pettit 1997, 2012).³⁸ The alien will need not be that of a superior. And, for that reason, the dominating will need not be that of another individual, but can instead be that of an institution or collective, with which questions of equality don't arise.³⁹

The objection to corruption, however, cannot be an objection to domination. If we are dominated by corrupt officials, then we are likewise dominated by non-corrupt officials, so long as they could, but won't, act as corrupt officials do. Moreover, once an official realizes that he *could* act corruptly, shouldn't he conclude that the damage is

³⁸ Similarly, some interpreters of Kant, such as Ripstein (2009), suggest that you are "dependent" when an alien, "unilateral" will could use or destroy your person or property.

³⁹ In other ways, however, domination is narrower than subordination. Domination is simply a matter of interference in choice, whereas subordination is a matter of power of other kinds, such as to withhold goods or change one's environment, as well as of authority and, as I discuss in section 10.3, "consideration."

already done? Why then should he feel any further qualms about *actually* acting corruptly?⁴⁰

Moreover, domination seems impossible, even in principle, to escape. What can possibly prevent one will from interfering, if not yet another will? But why aren't you then dominated by that other will—perhaps at the limit, some collective will, such as the Deep State or the People? The two features of domination that make it inescapable, however, are not shared by subordination. The first is that you are dominated when the alien will—say the state—so much as has a certain kind of *capacity*. By contrast, the other individuals who make up the state don't need to be somehow *incapacitated* in order for you not to be subordinate to them. It's enough that officials, *in fact*, exclude personalizing influences; that officials, *in fact*, favor no citizen over another (section 6); that officials, *in fact*, comply with decisions made in ways that all had equal opportunity to influence (section 10.1); that highest-level officials, *in fact*, conform to the standards of proper representation (section 10.2); and finally that we all, *in fact*, cede to one another, as a matter of right, equal opportunity for influence over political decisions (section 10.1). The second feature is that *collective* wills, no less than by individual wills, can dominate. By contrast, one can only be subordinated to another individual. Thus, one is not subordinated merely by being subject to a *collective* decision, so long as one does not thereby stand as an inferior to any other *individual*. For these reasons, among others, I believe that subordination better explains the intuitive anxiety that republicans are after, but ultimately mischaracterize, in their official definitions of domination (Kolodny forthcoming).

10. The diversity (and unity) of objections to money in politics

10.1. Equal influence

No doubt, corruption is one problem of money in politics. But there are others. Money in politics can disserve the public interest, in diverse ways. It can lead to bad decisions. It can lead to waste (not least the waste of politicians' time fundraising). And it can deprive aspirants of fair opportunity to hold elected office.⁴¹ What I want to highlight, however, are several other problems of money in politics, which are distinct from corruption, but which, like corruption, are also problems because of the threat that they pose to the ideal of a society of equals.

One such problem is that money in politics can undermine equal opportunity for informed, autonomous influence over political decisions. Equal influence (to give it a shorter label) is, like excluding personalizing influences, a way of moderating the threat of the subordination of some to others that the superior, final, and inescapable power and authority of the state would otherwise pose (Kolodny 2014b).

⁴⁰ There is also the problem of defining "arbitrary." Republicans sometimes define "arbitrary" interference as interference that does not serve the public interest. But, as we have seen, some corruption does serve the public interest, and some disservice of the public interest is not corrupt.

⁴¹ This is just a special case of a more general complaint, not special to politics: that no form of employment should be closed to one simply because one isn't rich (Beitz 1989: 174–5).

Equal influence requires different things at different locations in the democratic structure.

First, it requires that voters have equal access to the means to persuade other voters.⁴² While money in politics can lead to unequal access to the means of persuasion, such unequal access need not involve corruption. There need be no personalizing influence when a candidate passively benefits from lopsided “uncoordinated” expenditures, let alone when an ordinary voter is persuaded by Koch- or Soros-funded ads.⁴³

Equal influence also requires that officials comply with democratically made law about how their offices are to be discharged. If Bent defies such laws, which were made in a way that gave everyone equal opportunity for influence, then, whether or not he does so corruptly, he thereby deprives some people, “after the fact” as it were, of equal influence over those laws.

⁴² More precisely, it requires equal “judgment-independent” opportunity for influence. Unequal “judgment-dependent” influence—which depends simply on the influenced party’s free reflection on relevant reasons—is not in itself objectionable (Kolodny 2014b). Someone does not deprive you of equal opportunity simply because they find your case unpersuasive. This is because in finding your case unpersuasive, they are simply exercising an opportunity of the kind that you are claiming: namely, an opportunity to influence political decisions on the basis of *their* free reflection on relevant reasons.

The distinction between judgment-dependent and judgment-independent influence may explain why certain media endorsements do not constitute an objectionable inequality of influence. That is, it may provide a response to the view of the majority in *Citizens United* that the media exception of § 441b of BCRA was “all but an admission of the invalidity of the antidistortion rationale” (i.e., that “immense aggregations of wealth” by corporations might influence policy in ways that “have little or no correlation to the public’s support” (*Austin v. Michigan Chamber of Commerce*)). Suppose that people seek out the opinion page in the *Daily Grey*, based on an assessment of *Grey*’s reliability, judgment, congenial worldview, etc. on political questions. To the extent that the influence of *Grey*’s opinion page depends, in this way, on readers’ prior judgment that *Grey* has worthwhile things to say on political questions (Lowenstein 1985 846), that influence seems more “judgment dependent.” By contrast, to the extent that the *Red Network*’s influence results from *Red*’s interleaving sports and entertainment programming, or the only local news broadcast, with partisan editorials, it seems more “judgment independent.” Someone who merely wants to see the highlights from the game or tomorrow’s weather has not chosen to seek out what *Red* has to say on political matters.

⁴³ Some are tempted by the view that it’s not a problem when some citizens can spend more money to influence other voters, since voters in the end can, by their equal votes, ratify or reject those attempts at influence. A problem arises only when money is used to influence elected officials. That “short-circuits the democratic process” (Lessig 2011: 160–2; 2015: 58–60, 150–2; Thompson 1995: 113–115), since voters cannot ratify or reject those attempts at influence.

However, if attempts to influence officials are disclosed (and campaign contributions are for the most part disclosed), then voters can ratify or reject them by reelecting or replacing the official (Lowenstein 1995). And when the attempt to influence officials is via campaign contributions, voters can reject contributors’ attempts to influence officials simply by refusing to be influenced by what the contributions can buy (Cain 1995: 117).

Of course, an adherent of the view might reply that voters can’t just “refuse to be influenced” by what campaign contributions can buy. For example, suppose a few wealthy donors, by controlling the “money primary,” determine which candidates ever to come to voters’ attention. How can one “refuse to be influenced” by the fact that one is kept in ignorance of the alternatives? But this reply seems to concede the point: that it is a problem when some citizens can spend more money to influence other voters.

Finally, equal influence requires that, where officials enjoy legal discretion, some citizens don't have greater opportunity to influence how officials exercise that discretion, whether or not the influence is corrupt. Suppose, for example, that only a small circle of megadonors knew that a given legislator, Lexi, would not, if elected, act as most voters had been given to expect. Then, in the election itself, Lexi's circle, even merely as voters, had greater opportunity for *informed* influence than other citizens over the decisions that would be made by whichever legislator occupied the seat. Only they knew what they were getting.

10.2. *Proper representation*

Another problem of money in politics has to do not with citizens' relations to one another, but instead with their relations to officials. In her capacity as a legislator, Lexi can influence political decisions in a way in which no ordinary citizen can. Now perhaps this disparity in influence necessarily subordinates ordinary citizens to Lexi. But if not, it is presumably because certain "standards of proper representation" are met, so that Lexi stands as "agent" to the citizenry as collective "principal" (Kolodny 2014b, ms.). A further problem is that money in politics may violate these standards, again whether or not this happens by corruption.

One plausible standard is that the agent acts as the principal was given reason, at the time of selecting the agent, to expect that the agent would act. Another plausible standard is that the principal selects the agent, from an adequate range of alternatives, with periodic opportunities to renew or replace the agent. If wealth is a prerequisite for candidacy, however, then the range of alternatives from which representatives are selected may be too narrow.

10.3. *Favoritism*

One last objection is that money in politics involves favoritism (compare Thompson 1995 80–84). In particular, it's sometimes said, by way of defense, that money buys only access, not results. But buying access may already be favoritism. It's one thing for a legislator to stop taking calls from constituents. There are only so many hours in the day. But it's another to stop taking calls from some constituents, while still taking calls from donors. Even if being listened to has no chance of bringing about what the constituent seeks, being listened to (like Abel's having his application considered, section 6.3) matters, at least when others are being listened to.

Although favoritism, as we saw, is distinct from corruption, it also threatens the ideal of a society of equals. For relations of inferiority consist not only in asymmetries in power and authority, but also in disparities in "consideration": in certain kinds of favorable responses that people within a society receive, such as respect or concern (Kolodny 2014). What distinguishes the favorable responses constitutive of consideration, put somewhat circularly, is that disparities in such responses make for distinctions in rank or status in the broader society. What the state does for its citizens is just such a response. And this is why, as we saw in section 6.3, anti-favoritism, while not regulating everything that an agent does for others, regulates at least what the state does

for its citizens. When one private person looks after her friend's cat but not a stranger's, she does not somehow constitute the latter as second class. When the state provides relief aid for some citizens that it does not provide for others, by contrast, it does.⁴⁴

11. Conclusion

Many argue that, in order to understand the manifest deformation of American government by political money, and to vindicate the constitutionality of campaign finance reform, we need some "new" conception of corruption. This "new" conception must recognize that corruption takes forms that are not so much a matter of discrete, individual vice: forms that elude depiction in a Thomas Nast cartoon, by overstuffed pols with overstuffed pockets. Accordingly, Thompson (1995) seeks to articulate a new conception of "institutional" corruption, whereas Lessig (2011, 2015) seeks to articulate a new conception of "dependence" corruption.

These theorists are apt to be disappointed by Subordination by Personalization. Granted, it allows that corruption may not be conscious or deliberate. But it does say that corruption is a matter of how individual officials make decisions. And it in effect gives corruption a pass when there is no institutionally viable alternative.

In defense of Subordination by Personalization, I can only express my worry that, at least as a matter of conceptual hygiene, if not of legal strategy, these new, more expansive conceptions of corruption suppress relevant distinctions (compare Hasen 2016). While there are objections to the current regime of campaign finance, some of these are objections not to corruption, but instead, as we have said, to disserving the public interest, unequal influence, improper representation, or favoritism. The new conceptions, I worry, merely slap the label "corruption" on some independent objection.⁴⁵

Moreover, our discussion allows us to criticize the Court's reasoning, on two fronts. First, many of the measures that the Court has struck down or viewed with suspicion can perfectly well be justified as prophylactics against corruption, even on our narrow, railroad-baron-era conception. Why not worry that some legislators will act in office so as to attract or repay uncoordinated expenditures in ways that are hardly necessary for their offices to function well? That still counts as corruption, even after the concept has been sanded down.

Second, the Court's position, or at least the position that a lay observer could be forgiven for ascribing to the Court, seems unstable: namely, that corruption is the only pathology of the current regime of campaign finance serious enough to treat. For other

⁴⁴ Among the reasons for this are that the state, unlike a private person, has comprehensive reach; that it defines, in citizenship, a status more basic and less conditional than membership in most private associations; and that its power and authority are final and inescapable.

⁴⁵ For instance, Lessig (2011 8; 2015 4) all but identifies "dependence corruption," at least when applied to legislatures, with unresponsive policy (Dawood 2014): that is, policy that does not satisfy the policy preferences of the majority at any given time. But unresponsive policy can arise for a host of reasons, such as supermajority requirements, that have nothing to do with corruption on any recognizable use of the term. At any rate, as I discuss elsewhere, I don't think that responsiveness is an important value in itself (Kolodny 2014a, Kolodny ms. b).

pathologies of that regime, we have seen, include the ways that the regime threatens equal influence, proper representation, and anti-favoritism. And these, we have argued, are, like corruption, threats to the ideal of a society of equals. The point of equal influence and proper representation is to defuse the threat of subordination of some to others that the final, inescapable, and vastly superior power and authority of the state would otherwise pose. And the point of anti-favoritism is to prevent a disparity of consideration. If we approach the matter at the right philosophical depth, therefore, we should find it arbitrary to view only one of these threats to equality, namely corruption, as worth addressing.

This is, I conclude by venturing, further confirmation of the conjecture described at the start of this paper: namely, that at the root of many anxieties about the relation of the state to its citizens, or the relations of citizens to one another, lies an anxiety that none of us stands below, or for that matter above, any other.

Anderson, Elizabeth, 1997: “What is the Point of Equality?” *Ethics* XX: XX–XX.

Beitz, Charles, XXXX: *Political Equality* (XXXX).

Burke, Thomas, 2002: “Corruption Concepts and Federal Campaign Finance Law,” in Heidenheimer and Johnson (2002).

Cain, Bruce, 1995: “Moralism and Realism in Campaign Finance Reform,” *University of Chicago Legal Forum*: 111–140.

Dawood, Yasmin, 2014: “Classifying Corruption,” *Duke Journal of Constitutional Law & Public Policy*.

Dworkin, Ronald, 2002: *Sovereign Virtue* (Harvard).

Friedrich, Carl J., 1972: “Corruption Concepts in Historical Perspective,” in Heidenheimer and Johnson (2002).

Fukuyama, Francis, 2011: *The Origins of Political Order* (Farrar, Straus, and Giroux).

Gardiner, John, 1993: “Defining Corruption,” in Heidenheimer and Johnson (2002).

Hasen, Richard L., 2016: *Plutocrats United* (Yale).

Heidenheimer, Arnold J. and Michael Johnston, eds., 2002: *Political Corruption, Third Edition* (Transaction).

Hellman, Deborah, 2013: “Defining Corruption and Constitutionalizing Democracy,” 111 *Michigan Law Review* 1385.

Holmes, Leslie, 2015: *Corruption: A Very Short Introduction* (Oxford).

- Huntington, Samuel P., 1968: “Modernization and Corruption,” in Heidenheimer and Johnson (2002).
- van Klaveren, Jacob, 1957: “Corruption as a Historical Phenomenon,” in Heidenheimer and Johnson (2002).
- Leff, Nathaniel H., 1964: “Economic Development Through Bureaucratic Corruption,” in Heidenheimer and Johnston (2002).
- Kolodny, Niko, 2014a: “Rule Over None I: What Justifies Democracy?” *Philosophy and Public Affairs* 42: 195–229.
- Kolodny, Niko, 2014b: “Rule Over None II: Social Equality and the Justification of Democracy,” *Philosophy and Public Affairs* 42: 287–336.
- Kolodny, Niko, 2016: “Political Rule and Its Discontents,” *Oxford Studies in Political Philosophy* 2: 35–70.
- Kolodny, Niko, 2017: “What Makes Threats Wrong?” *Analytic Philosophy* 58: 87–118.
- Kolodny, Niko, forthcoming: “Being Under the Power of Others,” in Yiftah Elizar and Geneviève Rousselière, *Republicanism and Democracy* (Cambridge).
- Kolodny, Niko, ms. a: “Standing and the Sources of Liberalism”
- Kolodny, Niko, ms. b: “Democracy for Idealists”
- Lessig, Lawrence, 2011: *Republic, Lost: How Money Corrupts Congress—and a Plan to Stop It* (Hachette).
- Lessig, Lawrence, 2015: *Republic, Lost: The Corruption of Equality and the Steps to End It* (Hachette).
- Leys, Colin, 1965: “What is the Problem about Corruption?” in Heidenheimer and Johnson (2002).
- Lowenstein, Daniel H. 1985: “Political Bribery and the Intermediate Theory of Politics,” *UCLA Law Review* 32: 784–851.
- Lowenstein, Daniel H. 1995: “Campaign Contributions and Corruption: Comments on Strauss and Cain,” *University of Chicago Legal Forum*: 163–192.
- Lowenstein, Daniel H. 2004: “When Is a Campaign Contribution a Bribe?” in William C. Heffernan and John Kleinig, eds., *Private and Public Corruption* (Rowman and Littlefield).
- Moreau, Sophia, ms.: *The Many Faces of Inequality*.

- Nye, J. S., 1967: "Corruption and Political Development: A Cost-Benefit Analysis" in Heidenheimer and Johnston (2002).
- Pettit, Philip, 1997: *Republicanism* (Cambridge).
- Pettit, Philip, 2012: *On the People's Terms* (Cambridge).
- Philp, Mark, 1997: "Defining Political Corruption," *Political Studies* 45: 436–462.
- Philp, Mark, 2002: "Conceptualizing Political Corruption," in Heidenheimer and Johnson (2002).
- Raz, Joseph, 1997: "The Rule of Law and Its Virtue," *Law Quarterly Review*.
- Raz,
- Ripstein, Arthur, 2009: *Force and Freedom* (Harvard).
- Ryan, Alan, 2013: "Conceptions of Corruption, Its Causes, and Its Cure," *Social Research* 80: 977–992.
- Scanlon, T.M, forthcoming: *Why Does Inequality Matter?*
- Scheffler, Samuel, 2003: "What is Egalitarianism?" *Philosophy and Public Affairs* 31: 5–39.
- Strauss, David A. 1994: "Corruption, Equality, and Campaign Finance Reform," *Columbia Law Review* 1369.
- Strauss, David A. 1995: "What is the Goal of Campaign Finance Reform?" *University of Chicago Legal Forum* 141.
- Swart, Koenraad W. 1949: "The Sale of Public Offices," in Heidenheimer and Johnston (2002).
- Teachout, Zephyr, 2014: *Corruption in America* (Harvard).
- Thompson, Dennis F. 1995: *Ethics in Congress* (Brookings).
- Weber, Max 1956: *Economy and Society*.

Desire and Loving for Properties

Yongmin Han

Perhaps one of the deepest touchstones about how we come to love someone is that it's typically *selective*:

Liking Precedes Love: *Coming to love someone, at least if they are a non-family-member, often begins (though it doesn't always) with our liking (that is, being pleased by) them for certain properties of theirs: we wouldn't have come to love them unless we liked them.*¹

A tempting, indeed popular, way to vindicate such selectivity is to hold that love is a response to reasons; in particular, that it is a response to the beloved's qualities. To see what I mean, consider an analogy with emotions. According to cognitivists about emotions, what explains why we, e.g., come to admire someone in the wake of finding them admirable is that admiration is a response to, or is directly explained by, our finding them admirable. Similarly, in the case of love, it's tempting to think that what explains why we come to love someone in the wake of finding them attractive, is that love is a response to our perceiving their 'lovable' properties.

This paper takes an altogether different tack: I'll argue that we can, indeed should, vindicate love's selectivity by *denying* that love is a response to reasons. I'll do so by offering alternative explanations of data like Liking Precedes Love from the perspective of the desire-based view of love that I favor,

Humean Love: *To love someone is to strongly desire their well-being for its own sake, or 'intrinsically' (and intrinsic desires are not responses to reasons).*²

Now, my claim that Humean Love can explain love's selectivity might seem like a non-starter. After all, if love, as Humean Love has it, is oblivious to reasons, how could it be responsive to what the beloved's like? Part of the point of my explanations, then, is defensive: it's to show that we can explain selectivity even if we deny that love is a response to reasons. But my explanations will also form the basis for a *challenge* to views on which love is a response to reasons: Humean Love, I'll further argue, yields a better explanation than such views.

1. Introduction

Like most moral philosophers, I'm taking "love" as a theoretical term³ which denotes the attitude common to the varieties of love that people are said to have for their spouses, friends, and family members. This is an attitude we might pre-theoretically gloss as a 'concern for the beloved's well-being'.

Love's selectivity can seem to motivate the influential family of views that will be my foil:

Rationalism: *Love is a 'response to reasons'. That is, just as emotions and beliefs are (held by some to be) the sorts of states that can be both explained and justified by certain properties of their objects, so love is the sort of state that can be both explained and justified by objective evaluative properties of the beloved's.*⁶

Perhaps the most natural member of Rationalism to go for, given love's selectivity, is the *quality* view: Love is a response to qualities of the beloved's like their kindness or wit.⁴ For, as I've codified in *Liking Precedes Love*, our love for our partners and friends often seems to develop in the wake of liking their qualities.

To illustrate, consider how Elena Greco, the fictional writer of Elena Ferrante's Neapolitan novels, is first drawn to Lila Cerullo because of the latter's non-conformity and brilliant mind. Her love for Lila, furthermore, seems to develop in the wake of liking how, e.g., her adventures with Lila expand her horizons. What, the thought might go, explains how Elena's love develops in the wake of liking Lila, if not that her love is a response to (her assessments of) Lila's properties?

The problem with Rationalism, however, is that it can't seem to explain all of the data. To begin with, Rationalists can't seem to vindicate love's selectivity while accommodating the beloved's *irreplaceability*.⁵ We don't, after all, tend to 'trade up'. And while we sometimes stop loving someone after finding their properties unattractive, that's often not the case. Nor is it just about irreplaceability. Love doesn't, in general, seem reason responsive: we can't reason ourselves into (or out of) love. And we needn't, in coming to love someone, have perceived reasons for doing so.

Now, these worries have tended to leave most commentators unfazed: Rationalism continues to be the most popular approach, at least in print. But they at least raise the question: might *denying* Rationalism be a more promising way of vindicating selectivity while accommodating the ways in which love seems oblivious to reasons?

Not surprisingly, that's the route I prefer. In particular, I think that the desire-based view I favor allows us to explain all the data. This desire-based view is what I call Humean Love; it is the conjunction of these two claims:

Love as Desire: *To love someone is to desire their well-being strongly and intrinsically (i.e., for its own sake).*

Humean Desire: *Intrinsic desires, or desires that are had for their own sakes, are not responses to reasons.*

(Each of these claims is controversial.⁸ But I cannot take up the various objections here.)

Though simple,⁷ Humean Love already seems to capture some key facets of love. To see why, let's start by clarifying the type of desire that Humean Love says love is. We can distinguish among three types of desires: intrinsic, instrumental, and realizer.⁹ If one wants something (at least partially) for its own sake, call that an *intrinsic* desire.¹⁰ If one wants it merely as a means to some intrinsically desired end, call that an *instrumental* desire. And if one wants it merely because it's a specific realization of some intrinsically desired end, call that a *realizer* desire.¹¹

By holding that love is an *intrinsic* desire, then, Humean Love codifies the way love seems *disinterested* or *selfless*. Furthermore, it gets at the sorts of *motivational, emotional, and cognitive dispositions* we associate with lovers. For if one intrinsically desires some entity's good, one will be disposed, *inter alia*, to act to promote their good, to be pleased

when they fare well and displeased when they don't, and to notice and attend to things associated with them and their interests more than people who don't love them.¹²

Let me conclude by elaborating on the difference between intrinsic and non-intrinsic desires. Consider Khairul, a researcher who initially wants emerging technologies to be as free from regulation as possible, because that's what preliminary research suggests would best encourage economic growth. But, after doing more research, he comes to believe that more regulations will actually aid, rather than hamper, economic growth; and as his beliefs change, so, apparently, do his desires: he now wants these technologies to be governed with a more heavy hand.

Khairul's example may seem to suggest, *contra* Humean Desire, that some intrinsic desires are reason responsive (or, more strongly, that it is not characteristic of intrinsic desires that they are oblivious to reasons). But that would be the wrong conclusion to draw. The reason why Khairul's desire changes with his beliefs is that it's a *non*-intrinsic desire. It's an instrumental desire that's brought about, and causally sustained, by the interaction of (e.g.) his intrinsic desire for his nation's economic growth with his belief that imposing fewer regulations is a means to economic growth.¹³ Its existence *depends* on the prior existence of that intrinsic desire and belief.

By contrast, intrinsic desires are more resistant to changes in our relevant beliefs and desires. For example, the intrinsic desire most of us have for sweet foods will tend to survive changes in our other desires and beliefs. It wouldn't vanish upon our becoming convinced that sweet foods are undesirable (e.g., because unhealthy); or at least, it wouldn't tend to do so *in the way that* our admiration might were we to find the person unadmirable.

2. Coming to want something can seem like coming to love someone

Claiming that Humean Love can vindicate Liking Precedes Love, however, might seem absurd. In general, there are two ways to explain why *A* often precedes *B*: either *A* explains *B*, or there is some third factor, *C*, which explains both *A* and *B*. Rationalists, as we've seen, can in principle go for the first option: Elena's liking Lila (which they might ground in terms of perceiving reasons) explains her coming to love Lila. Anti-Rationalists like Humean Love must instead go for a third factor explanation. But it's not obvious what the third factor might be.

Advancing such a third factor explanation is the task of this and the next section. First I'll observe that other intrinsic desires can similarly appear selective and yet oblivious to reasons. Then, building upon Nomy Arpaly and Tim Schroeder (2014)'s discussion, I'll develop a Humean account of why, in general, coming to desire something often begins with our liking it, an account that I'll extend in the next section to desiring someone's well-being.

To see why other desires seem to pattern like love, consider how

As a child, Rebecca first starts wanting to read fiction because reading and discussing fiction is one of the things that her friends often did. But the more fiction she reads, the more she relishes the way reading fiction gives her insight into the human condition (though, of course, she wouldn't have put it that way back then). Before long, she comes to

be interested in reading, or appreciating, fiction for its own sake. For instance, she continues wanting to read fiction and think about it, even when she no longer has opportunities to talk about books with her friends.

Just as Elena's coming to love Lila begins with her liking Lila's properties, so Rebecca's coming to want to read fiction begins with her liking fiction — in particular, with her liking how it would allow her to spend more time with her friends, as well as its humanistic insights. In short, as Arpaly and Schroeder (2014) point out, the process of coming to desire something intrinsically patterns like that of coming to love someone in that

Liking Precedes Desire: *Coming to desire something intrinsically often begins with one's liking it.*

Furthermore, the objects of desire also seem oblivious to reasons. For instance, just as, having come to love Lila, Elena will continue loving her even when Lila seems to her to no longer be interested in intellectual pursuits, so, having come to want to read fiction for its own sake, Rebecca continues wanting to do so even when she no longer gets to have book discussions with her friends.

So, a natural strategy is for Humean Love to explain love's selectivity by deploying, and extending, its account of *desire's* selectivity. That is, perhaps whatever explains why Rebecca's coming to desire to read fiction begins with her liking it can be extended to explain why Elena's coming to love Lila begins with her liking Lila.

2.1. Explaining Liking Precedes Desire

This is the strategy that I will pursue. To see how this might work, consider the process of coming to want something for its own sake. This process has two stages: although it ends with our wanting the thing for its own sake, it usually begins with our wanting the thing only for the sake of our other ends.¹⁴

Let me walk through this process with Rebecca's example. We've seen how part of why Rebecca initially wants to read fiction is that she wants to spend time with her friends, and that she wants to understand the human condition. That is, she begins with both a *realizer* and *instrumental* desire to read fiction. She begins wanting to do it both as a way of spending time with her friends, and as a means to obtaining humanistic insights. And she begins with these non-intrinsic desires, because she has the relevant antecedent beliefs and intrinsic desires: she believes that reading (and talking about) fiction is a way in which she and her friends spend time and she intrinsically desires to spend time with them; and she sees that reading is a means to understanding the human condition and intrinsically desires to understand the human condition.

I flag this distinction between beginning with instrumental desires and beginning with realizer desires, only to note that it isn't crucial here. Before explaining why, however, let me emphasize that this sort of transition from non-intrinsic desires to intrinsic desire is familiar. Frida begins wanting to eat Thai cuisine merely because her friends really liked Thai restaurants. Her initial desire, in other words, is a realizer one that's produced by her intrinsic desire to hang out with her friends. But after numerous such meals with her friends, she finds herself hankering for Thai food, even when other options are available. And the phenomenon, if anything, is even more familiar in the case of instrumental desires.

We might start off wanting to celebrate the new year only in order to spend time with family and friends. But, after years of such festivities, we often come to want to ring in the new year for its own sake.

This sort of transition from non-intrinsic to intrinsic desire, then, is familiar. Although the properties of the desired thing largely cease to be as relevant by the end of the process, they are often relevant at the start, since we begin by *non*-intrinsically desiring the thing. But it doesn't matter, at the start, *how* the properties are relevant; all that matters is *that* they are relevant.

We've seen how the first stage of the process usually comes about. But this causal explanation already gives us the materials for a third factor explanation of Liking Precedes Desire. The causal explanation, as we've seen, implies that Rebecca begins with desires that will be satisfied by her appreciating fiction, such as the desire for humanistic insights. But if so, it's no surprise that she also begins by *liking* fiction-appreciation. For these desires will tend to cause her to be pleased by, and thereby like, fiction-appreciation. (Again, I am operationalizing liking something as being pleased by it.) These desires, that is, will tend to cause Rebecca to be pleased by reading or appreciating fiction, since they will tend to cause her to feel pleasure when doing so enables her, e.g., to obtain humanistic insights. For we tend to feel pleasure and other positive emotions, to the extent that we seem to get what we want.

On this account, then, what explains why coming to want something begins with our liking its properties — with the thing's properties seeming relevant — isn't that intrinsic desires are reason-responsive. It isn't that our perceiving its attractive properties (or other `reasons for wanting it') *directly explains* our coming to want it in the way that finding someone admirable (on cognitivism about emotions) directly explains our coming to admire them. Rather, it's that there's often a *third factor* — our having other intrinsic desires that seem to be satisfied by the thing — which (partially) explains both our coming to want it and our liking it.

2.2. Complications

2.2.1. Why not go for a Rationalist explanation?

I've presented a (merely) *causal* account of why Rebecca's coming to want to appreciate fiction for its own sake begins with her liking it. But why not instead go for a Rationalist explanation: why not say that her desire to appreciate fiction grows in the wake of liking it, because it's explained, indeed justified by, her recognition of fiction appreciation's qualities?

Response: we should prefer my causal account because it's *simpler*. For Rationalists also need the sorts of resources that Humean Love invokes. To begin with, they seem to need such resources to explain how conceptually unsophisticated children and animals acquire intrinsic desires. Furthermore, Rationalists should agree that *some* cases of intrinsic desire acquisition come about from mere psychological causes, rather than from appreciation of reasons (e.g., the new year celebration case, where more *circumstantial* factors seem to be what's at play). So, Rationalists must hold that it's a merely causal account like mine in some cases, but reasons in others. But insofar as my causal account can *also* explain those

cases where Rationalists are inclined to appeal to reasons, my account makes for a simpler explanation of the data.

2.2.2. Rejoinder

I've said nothing about what explains how those initial non-intrinsic desires transform into an intrinsic one. One may therefore worry that this leaves open the possibility that perceiving reasons plays some role in explaining the second stage of the process. So, it's worth noting that plausible Humean explanations are available also for the second stage. We might hold, for instance, that Rebecca's non-intrinsic desires become intrinsic because she's *associated* fiction appreciation with things that she antecedently desires, such as understanding the human condition.¹⁵

3. Applying this account to love

Suppose we take love, as Humean Love has it, to be a strong, intrinsic desire for the beloved's well-being. Extending the third factor explanation of desire's selectivity to love is, in broad outline, straightforward. On this account, it's not that Elena's liking (or perceiving Lila's 'lovable' properties) directly explains her coming to love Lila in the same way that our finding someone admirable might directly explain our coming to admire them. Instead, as with Rebecca, it's that Elena's coming to desire Lila's good for its own sake begins with (because partially explained by) her having other sorts of intrinsic desires; in particular, with desires that are satisfied by being around Lila, such as the desire to be around brilliant people. These other desires are what cause her to be pleased by qualities of Lila's like her brilliance, before coming to her (Liking Precedes Love).

This schematic account, however, may seem only to raise more question. How, we might wonder, does Elena get from liking Lila for her brilliance, or wanting to be around her because she's brilliant, to wanting Lila's well-being? Why, that is, does Elena not simply come to intrinsically want to be around Lila — why does she also come to want, of all things, Lila's *good*?

Here's a proposal for one (though not necessarily the only) way it could happen. Just as Rebecca begins by non-intrinsically desiring to read fiction, so Elena begins by non-intrinsically desiring Lila's good. In particular, Elena begins with a realizer desire for Lila's good, because (i) she begins by liking, and believing that she likes, Lila and (ii) we intrinsically desire the good of that which we like (or, more precisely, that which cause us to experience pleasure). That is, most of us, I'll argue, have the general intrinsic desire for the good of that which we like (are pleased by). (Narcissists don't, but that's because narcissists don't care about anyone but themselves.) So, in the same way that the general desire most of us have to relieve others' suffering will generate the realizer desire to relieve a specific person's good when we see that they are suffering, Elena's general desire for the good of that she's pleased by will generate the realizer desire for Lila's good when (it seems to her that) she's pleased by Lila.

As with any other intrinsic desire, we can ascertain whether we have this general desire by investigating whether we have the dispositions that we would expect it to ground. Because these dispositions are the same in kind as those that love gives rise to, it'll simplify matters to consider only entities we don't already love. Think, for instance, about acquaintances we

have enjoyed interacting with but don't love, such as long lost schoolmates. It's natural, when we recollect the fun times we've spent in their company, for us to then come to manifest the dispositions characteristic of the desire for their well-being: to find ourselves wondering, say, how they are doing, and to hope that they are flourishing; or to become motivated to help them, if we happen to be in a position to greatly benefit them. What seems to explain such data, is that we have the intrinsic desire for the good of those we are pleased by, and that desire generates the realizer desire for the good of that person when it seems to us that we're pleased by them.

Moving now to more general patterns, consider how, even among those colleagues we don't already love, we tend to worry and look out more for those we like than those we don't like. Similarly, we're more greatly pleased by the flourishing of those sports teams we like than by those we don't, even if we aren't fans of any one of them. A natural explanation of this systematic difference, again, is that we intrinsically desire the good of that which we are pleased by.

I've just advanced a proposal for how the desire for the beloved-to-be's well-being might enter the picture. But this proposal, as it turns out, yields a *further* explanation of Liking Precedes Love. For, on this proposal, it isn't just that liking can precede coming to love someone because there is a third factor which partially explains both coming to love them and being pleased by them. It's also that liking (being pleased by) the person is, at least sometimes, part of what causally explains the first stage of coming to love them (i.e., our coming to non-intrinsically desire their well-being).

This claim, to be clear, doesn't commit me to Rationalism. For, even on this proposal, liking still doesn't bring about the (strong and) intrinsic desire for the beloved's well-being in the way that finding someone admirable might bring about admiration. All that liking can be a cause of is the *non*-intrinsic desire for the beloved's good — a state that the Humean would agree is 'reason-responsive' (in the sense of being responsive to changes to beliefs). Consequently, my account can accommodate the ways in which love seems oblivious to reasons.

Finally, this account allows the anti-Rationalist to go on the offensive. One criticism that we'd be able to press against Rationalists, of course, is that they can't explain love's selectivity while accommodating the ways in which it seems oblivious to reasons. But even if Rationalism, as its more sophisticated proponents would argue, can do so, we should still prefer Humean Love because it's *simpler*. For Rationalists also need the sorts of resources my Humean account invokes. They need, for instance, to posit the presence of an intrinsic desire for the good of that we've been pleased by, in order to explain the data I've adduced. And they need my Humean account, in order to explain cases of intrinsic desire acquisition that stem from mere psychological causes, as well as how the conceptually unsophisticated acquire intrinsic desires. It's just that they think that we need *additional* resources — reasons — to explain how we (more sophisticated creatures) sometimes come to love someone.

4. Conclusion

Now, one might worry, again, that we might need reasons to explain how these non-intrinsic desire(s) for Lila's good become intrinsic.¹⁶ But, as with Rebecca, it's natural to

take the cause to be Elena's associating Lila with desired situations such as being around brilliant people and the expansion of her horizons.

Let's take stock. I've developed and defended an unconventional, indeed overlooked, approach to vindicating love's selectivity (while accommodating the ways in which love doesn't seem reason-responsive), one which does so by denying that love is a response to reasons. This approach to vindicating love's selectivity will, of course, be helpful for anti-Rationalists. But it also has a critical upshot. For we now have a *new* reason to deny Rationalism. It isn't just that Rationalists may not be able to accommodate the ways in which love doesn't seem reason-responsive — it's also that, *even if* Rationalists can accommodate those data, denying Rationalism nevertheless furnishes us with *better* explanations.

References

- Abramson, Kate, and Adam Leite. 2011. "Love as a Reactive Emotion." *The Philosophical Quarterly* 61 (245): 673–699. [🔗](#)
- Arpaly, Nomy, and Timothy Schroeder. 2014. *In Praise of Desire*. Oxford University Press. [🔗](#)
- Bagley, Benjamin. 2015. "Loving Someone in Particular." *Ethics* 125 (2): 477–507. [🔗](#)
- Delaney, Neil. 1996. "Romantic Love and Loving Commitment: Articulating a Modern Ideal." *American Philosophical Quarterly* 33 (4): 339–356. [🔗](#)
- Ebels-Duggan, Kyla. 2008. "Against Beneficence: A Normative Account of Love." *Ethics*. [🔗](#)
- Frankfurt, Harry G. 1999. "Necessity, Volition, and Love." In . Cambridge University Press. [🔗](#)
- Frankfurt, Harry G. 2004. *The Reasons of Love*. Cambridge University Press. [🔗](#)
- Green, O. H. 1997. "Is Love an Emotion?" In *Love Analyzed*. Westview Press. [🔗](#)
- Jollimore, Troy. 2011. *Love's Vision*. Princeton University Press. [🔗](#)
- Keller, Simon. 2000. "How Do I Love Thee? Let Me Count The Properties." *American Philosophical Quarterly* 37 (2): 163–173. [🔗](#)
- Kolodny, Niko. 2003. "Love As Valuing A Relationship." *The Philosophical Review* 112 (2): 135–189. [🔗](#)
- Martin, Adrienne M. 2015. "Love, Incorporated." *Ethical Theory and Moral Practice* 18: 691–702. [🔗](#)
- Mill, John Stuart. 1863. *Utilitarianism*. London: Parker, Son and Bourn. [🔗](#)
- Schroeder, Timothy. 2004. *Three Faces of Desire*. Oxford University Press. [🔗](#)
- Schroeder, Timothy. 2015. "Desire." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2015. <http://plato.stanford.edu/archives/sum2015/entries/desire/>. [🔗](#)
- Setiya, Kieran. 2014. "Love and the Value of a Life." *Philosophical Review* 123 (3): 251–280. [🔗](#)
- Sinhababu, Neil. 2009. "The Humean Theory of Motivation Reformulated and Defended." *Philosophical Review* 118 (4): 465–500. [🔗](#)
- Smuts, Aaron. 2014. "Normative Reasons for Love, Part I." *Philosophy Compass* 9 (8): 518–526. doi:[10.1111/phc3.12168](https://doi.org/10.1111/phc3.12168). [🔗](#)
- Velleman, J. David. 1999. "Love as a Moral Emotion." *Ethics* 109 (2): 338–374. [🔗](#)

Wolf, Susan. 2002. “The True, The Good, and the Lovable: Frankfurt’s Avoidance of Objectivity.” In *The Contours of Agency: Essays on Themes from Harry Frankfurt*, 227–244. The MIT Press. [P](#)

¹Jollimore (2011), xiii, 20; Wolf (2002), 236; Kolodny (2003), 171. [↵](#)

²This specific view is formulated and briefly defended in Arpaly and Schroeder (2014). Other desire-based theorists include Frankfurt (2004), Smuts (2014), and Green (1997). [↵](#)

³This usage, while theoretically regimented, *is* one of the ways in which we can use “love” — and, for that matter, “care” — in ordinary English. I call it a theoretical term, however, to stress that the object of inquiry isn’t how “love” (or “care”) is used in ordinary English, but rather the nature of the attitude I’ve picked out. [↵](#)

⁴E.g.: Delaney (1996), Keller (2000), Jollimore (2011), Abramson and Leite (2011), Bagley (2015). [↵](#)

⁵More carefully: if we try to vindicate love’s selectivity by holding that love is a response to the beloved’s attractive properties, as the *quality view* does, we wouldn’t be able to accommodate these ways in which it *doesn’t* seem selective. Conversely, if we try to make room for the latter sorts of data by moving to other specifications of Rationalism, such as views on which love is a response to our relationship with the beloved, we wouldn’t be able to rationally vindicate selectivity. [↵](#)

⁶To sample some views, Rationalists hold, variously, that love is a response to the beloved’s *qualities* (Jollimore (2011)); to their *capacity for rational, self-reflective choice* (Velleman (1999)); to their *humanity* (Setiya (2014)); to their *moral virtues* (Abramson and Leite (2011)); or to the *relationship or history* they share with the lover (Kolodny (2003)). [↵](#)

⁷It is simpler than perhaps the most prominent of desire-based views, Frankfurt (1999)’s, which further requires higher-order conative endorsement of the first-order desire. [↵](#)

⁸For objections to Love as Desire, see, e.g., Velleman (1999), Bagley (2015) and Ebels-Duggan (2008). [↵](#)

⁹This discussion is meant to be more terminological than substantive. So, for example, the difference between I’m calling *intrinsic* and *non-intrinsic* desires just is the difference between what Kolodny (2003) calls *non-instrumental* and *instrumental* valuation. [↵](#)

¹⁰As Schroeder (2015) notes, one can also instrumentally desire that which one intrinsically desires. I may have an intrinsic desire for my friend’s well-being; but I can also desire it because I would be miserable if she were to suffer, and I desire not to be miserable. [↵](#)

¹¹Schroeder (2004). [↵](#)

¹²See Arpaly and Schroeder (2014), Schroeder (2004), and Sinhababu (2009) for discussions of desire’s effects. [↵](#)

¹³Arpaly and Schroeder (2014), 8-9. ↵

¹⁴Schroeder (2004), 147; Arpaly and Schroeder (2014), 101-2. ↵

¹⁵Mill (1863),26, Schroeder (2004), 147. ↵

¹⁶Martin (2015)'s view of love can be read this way. ↵

Mutual Recognition, Friendship, and the Moral Relationship

Richard Healey

Richard Healey is currently a Postdoctoral Fellow at the Centre de Recherche en Éthique and McGill University in Montréal. His research focuses on the power of consent, moral rights and duties, and interpersonal moral relationships. He is currently working on questions about the moral basis of our power of consent, and the nature and moral significance of "collective consent" in political contexts.

Moral agents are often thought to stand, or to be capable of standing, in a kind of valuable relationship with one another. The relationship in question is often characterized by way of an analogy with special relationships such as friendship, at least when things go well.⁴⁶ Yet this analogy can seem strained. Friendships are historical and on-going relationships between individuals that are quite distinct from otherwise decent moral relations between people.⁴⁷

My aim in this paper is to propose a novel way of making sense of the analogy between friendship and the “moral relationship.” I do so by developing a conception of *mutual recognition* between persons.⁴⁸ Specifically, I argue that a concrete form of inter-subjective recognition endows our interpersonal relations with a distinct form of value. This form of recognition is well exemplified by friendships, as well as other forms of special relationship like marriage, and collegial relationships. But these examples also help us to see that mutual recognition is a valuable constituent of our interactions and relations with those with whom we do not share special relationships – with those with whom we have nothing more than a basic moral relationship, we might say.

In developing this account, however, I am also pursuing another more general aim. When thinking about the nature and grounds of moral and political norms, appeals to valuable relationships are commonplace. I suspect, however, that such appeals – in

⁴⁶ See, for example, Christine Korsgaard, “Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations,” *Philosophical Perspectives* 6 (1992); T.M. Scanlon, *What We Owe To Each Other* (Cambridge, Mass.: Harvard University Press, 1998); T. M. Scanlon, *Moral Dimensions: Permissibility, Meaning, Blame* (Cambridge, Mass.: Harvard University Press, 2008), esp. pp. 139-41; R. Jay Wallace, “The Deontic Structure of Morality,” in *Thinking about Reasons: Themes from the Philosophy of Jonathan Dancy*, eds. David Bakhurst, Brad Hooker, and Margaret Olivia Little (Oxford: Oxford University Press, 2013).

⁴⁷ See, e.g., R. Jay Wallace, “Dispassionate Opprobrium: On Blame and the Reactive Sentiments,” in *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, eds. R. Jay Wallace, Rahul Kumar, and Samuel Freeman (Oxford: Oxford University Press, 2011), pp. 358-60.

⁴⁸ This account of mutual recognition is distinct from the usage that Scanlon makes of this term in *What We Owe To Each Other*. I explain the differences between his account and my own, and my reason for employing the same term, below.

discussions of rights, interpersonal justification, blame, social justice, and democratic theory, to give just some contemporary examples⁴⁹ – are appeals to a family of related but importantly distinct ideas. Yet the ways in which these ideas are distinct is rarely made clear. Thus, in developing the following account of mutual recognition, I hope to make explicit a particular kind of value that is realized within our personal relationships, and in so doing, to contribute to an understanding of the different ways in which valuable interpersonal relationships figure in our moral and political thought. In particular, I hope to emphasize the centrality of a certain form of concrete inter-subjective recognition to our moral lives.

My outline of the idea of mutual recognition begins with the observation that the value of many of our interactions and relationships does not only depend upon the way in which individuals act. It also depends upon the way in which agents deliberate about how to act: the way in which individuals take one another's interests, needs, and choices into account when engaged in practical reasoning.⁵⁰ As Niko Kolodny puts it, "It matters to us not only that...people and things fare well in nature: that they escape harm, flourish, and so on. It matters to us also that they are properly regarded by others."⁵¹ One context in which this is particularly clear is the context of our special relationships. We believe that a friendship between A and B is only a valuable or flourishing example of a friendship, for instance, if A and B *relate* to one another as friends: if A and B are each disposed to deliberate about matters that fall within the parameters of their relationship in a certain fashion. A and B ought, for example, to recognise special reasons to provide emotional encouragement and support, to keep one another's confidences, to help in the pursuit of one another's projects, and so on. If A fails to recognise and respond appropriately to these reasons then, however she may act, she fails to relate to B as a friend, and this fact undermines the value of her relationship with B.

Call this the *deliberative requirement* on the value of relationships. While perhaps more familiar in the case of relationships such as friendship and marriage, a deliberative requirement plausibly holds outside of our special relationships. Just as we are concerned with the ways in which our friends and partners relate to us, we are also concerned with the way in which others relate to us – be they perfect strangers, baristas, co-citizens, or whatever – and not only with whether they act in ways that happen to serve or set back our morally significant interests. We will, for example, seriously resent someone who drives recklessly whilst drunk even if they do cause anyone any harm, because this behaviour

⁴⁹ In addition to the references in footnote 1, see, Elizabeth S. Anderson, "What is the Point of Equality?" *Ethics* 109, no. 2 (1999); Niko Kolodny, "Rule Over None II: Social Equality and the Justification of Democracy," *Philosophy and Public Affairs* 42, no. 4 (2014); Daniel Viehoff, "Democratic Equality and Political Authority," *Philosophy and Public Affairs* 42, no. 4 (2014); and Ralph Wedgwood, "Two Grades of Non-consequentialism" *Criminal Law and Philosophy* (2014), esp. § 5.

⁵⁰ For the sake of brevity, I henceforth refer only to the way in which one another's "interests" figure in our practical deliberations. Furthermore, I assume throughout that our interests, needs, and choices give other agents reasons, without specifying in any detail the precise relationship between the two.

⁵¹ Niko Kolodny, "Which Relationships Justify Partiality? The Case of Parents and Children," *Philosophy and Public Affairs* 38, no. 1 (2010), pp. 43-4.

manifests blatant disregard for others.⁵² Likewise, a Tyrant who oppresses a minority group is blameworthy not only because her actions leave the oppressed with an insufficient share of basic goods (e.g. food, shelter, and liberty), but also because in deliberating about the distribution of those goods she fails to give the interests of the oppressed group sufficient weight. Independently of the consequences, we might say, those who relate to others in such a way fail to treat them with the moral respect they deserve, by giving appropriate weight to their morally significant interests. This fact impairs the value of the relationship or interaction between them.

Plausibly, the deliberative requirement flows from our capacity to recognize and respond to reasons. That we can recognize and respond appropriately to the reasons we have within a particular relationship means that the way in which we deliberate affects the meaning and value of that relationship. But this capacity also results, as Rahul Kumar has observed, “in a distinct kind of vulnerability, a vulnerability to what another’s reasons, or reasoning...says about oneself, or one’s value as a person.”⁵³ What is more, this vulnerability informs many of the relationships and interactions that we engage in. Because we are capable of making judgements about how others do in fact deliberate, it matters to us that we engage in certain kinds of activity, and certain kinds of relationship, with others who (we judge to) relate to us in an appropriate manner for an activity or relationship of this kind.⁵⁴ Likewise, insofar as others are capable of making judgements about the reasons that we respond to in our dealings with them, they only wish to interact with us when (they judge) we respond appropriately to certain reasons, e.g., the requirements of friendship, or morality.

What examples such as friendship and marriage further help to show is that our capacity to recognize and respond to reasons does not *only* make us vulnerable to others in a way that is not true of non self-conscious agents. In addition, it opens up the possibility of a distinct form of valuable relationship that would otherwise be impossible: relationships in which two individuals mutually recognize that they each relate to one another in a way that is fitting for a valuable relationship of this kind, given the reasons they each have in the context of this relationship.

I can now say, more precisely, that two individuals stand in a relationship of mutual recognition when: (i) they both meet the deliberative requirement that holds within their relationship, and (ii) they share a common belief that they each meet this deliberative requirement.⁵⁵ Where these conditions hold, two individuals will mutually understand

⁵² Rahul Kumar, “Who Can Be Wronged?” *Philosophy and Public Affairs* 31, no. 2 (2003), p.103.

⁵³ *Ibid.*, p. 107.

⁵⁴ Cf. Rousseau: “As soon as men had begun to appreciate one another and the idea of consideration had taken shape in their mind... any intentional wrong became an affront because, together with the harm resulting from the injury, the offended party saw in it contempt for his person, often more unbearable than the harm itself.” Jean-Jacques Rousseau, “Discourse on the Origin and Foundations of Inequality Among Men,” in *The Discourses and Other Early Political Writings*, ed. Victor Gourevitch (Cambridge: Cambridge University Press, 1997), p. 166.

⁵⁵ For current purposes, A and B can be said to have a common belief that they each satisfy the deliberative requirement (DR) when:

(i) A and B believe that the other satisfies DR

themselves and one another as, for example, party to a friendship or marriage. They will therefore mutually recognize that they respond appropriately to one another's interests. This fact is, I want to claim, constitutive of the value of many forms of interaction and relationship. The value of friendship is, for example, partly constituted by the fact that A and B mutually recognize that they each deliberate in accordance with the normative requirements of friendship. The activities and pursuits of friendship are endowed with a special significance precisely if and because A and B have a common belief that they are friends.

Moreover, I think that reflection upon the phenomenology of our interactions and relations with one another supports the claim that mutual recognition is not only constitutive of the value of special relationships, but also of our general moral relations with one another.⁵⁶ As I have already said, we are concerned, as individuals, with whether others relate to us in a morally appropriate manner. Why not think that we also value relating to one another in a way that allows us to mutually recognize that we give one another's morally significant interests appropriate consideration? Take, for example, a one-night stand between two strangers, X and Y. I take it that there is no interesting (or, at least, very thick) sense in which X and Y stand in a special relationship just because they have sex. Nonetheless, I am confident that we will hold that the value of their sexual interaction depends, in part, on whether X and Y stand in a relationship of moral mutual recognition: whether they have a common belief that they each give one another's interests the weight that, morally speaking, they ought to give them. Similarly, my interaction with my surgeon (who, in reality, I might only meet for a five-minute consultation before I am taken away and anesthetized) is more valuable if it is characterized by mutual recognition between us, as opposed to say, mutual distrust or suspicion. And as this case suggests, the focus of our common belief can be (to some extent) asymmetrical. What is most significant in this case is that the surgeon and I share a common belief that the *surgeon* will relate to *me* appropriately by trying her best to restore me to full health, and so on.

To reinforce this idea, consider a case in which moral mutual recognition is lacking. Imagine that I arrive in country C on a business trip. A local guide informs me that the inhabitants of C do not regard me as having the same moral rights as them, or indeed as having any moral standing whatsoever. They believe that since I am not from C I fall outside of the moral community, and so am not worthy of (moral) concern. This discovery would significantly affect me and my thoughts about the relations I am able to have with the inhabitants of C.⁵⁷ This would be true even if I could be guaranteed that my basic rights would be respected for other (say, financial) reasons, and even if I achieve exactly the same ends as I otherwise would have (e.g. signing the contract). The fact that the inhabitants of C do not regard me as having morally significant interests, and my awareness of this fact, will impair the value of my relations with them in part because we cannot stand in the valuable relation of mutual recognition.

(ii) A and B both believe that the other believes that they satisfy DR.

⁵⁶ See Scanlon, *Moral Dimensions*, pp., 139-41.

⁵⁷ Scanlon, *What We Owe To Each Other*, p. 76.

With the general account of mutual recognition in place it is worth expanding on several features of the account. First, on this view, mutual recognition is a formal feature of relationships. That is to say, for any kind of relationship (R_x), in which a set of reasons and duties (D_y) apply to the participants (P1 and P2) of that relationship, P1 and P2 realize a relation of mutual recognition when they (i) both reason in accordance with D_y (and so meet the deliberative requirement), and (ii) P1 and P2 share a common belief that they each reason in accordance with D_y . As preceding examples have highlighted, the content of D_y will vary depending on the kind of relationship that P1 and P2 stand in. If P1 and P2 are friends then they will realize a relation of mutual recognition if and only if they both reason in accordance with the special reasons and duties of friendship, and they share a common belief that this is true. If P1 and P2 share no special relationship, by contrast, they realize a relation of mutual recognition just when they share a true common belief that they reason in accordance with (say) the moral reasons that apply to them with regard to one another. We can designate these different kinds of relationships of mutual recognition using subscripts. In the first case P1 and P2 stand in a relation of $MR_{\text{FRIENDSHIP}}$; in the second case they stand in a relation of MR_{MORAL} .

A second point is related. Because mutual recognition is no more than a formal feature of relationships, there is no guarantee that when two agents stand in a relationship of mutual recognition, their relationship is a non-instrumentally valuable relationship. P1 and P2 could stand, for example, in a relationship of MR_{MAFIA} or MR_{NAZISM} . It is a difficult further question, which I cannot resolve here, whether the fact that P1 and P2 stand in a relation of mutual recognition always endows a relationship with *some* non-instrumental value, even if the relationship is otherwise valueless.⁵⁸ However, I am inclined to say that this is false. On this view, mutual recognition only endows a relationship with a special form of value when the reasons and duties that P1 and P2 recognize as constitutive of their relationship track other non-instrumental values and reasons.⁵⁹ But there is room for disagreement here: Some might be inclined to think that the fact that we share a common belief that we are bound by a set of normative requirements suffices always to give our relationship some value.⁶⁰

It may also be useful to explicitly contrast my use of the term “mutual recognition” with T. M. Scanlon’s usage in *What We Owe To Each Other*. Scanlon argues that when we deliberate and act in accordance with principles that others could not reasonably reject, we stand in a relation of “mutual recognition” with them, a relation that is “appealing in itself – worth seeking for its own sake.”⁶¹ On Scanlon’s usage, at least as I understand it,

⁵⁸ It could, of course, always have instrumental value.

⁵⁹ Cf. Joseph Raz’s argument that personal autonomy is only valuable when exercised in pursuit of genuinely valuable projects and relationships in his, *The Morality of Freedom* (Oxford: Clarendon Press, 1986), Ch. 13.

⁶⁰ Think, for example, of those who hold we are capable of making valid immoral promises (e.g. a promise to do my part in our plot to blow up a school). Indeed, one might even hold that such promises are binding partly in virtue of the value of our mutually recognizing that I am bound to you by my voluntarily taking on this obligation. This is consistent with thinking that mutual recognition is *more* valuable in relationships where the requirements we recognize as binding track genuine non-instrumental values.

⁶¹ Scanlon, *What We Owe To Each Other*, p. 162. See also his *Moral Dimensions*, esp. pp. 139-41; and R. Jay Wallace, “The Deontic Structure of Morality,” in *Thinking about Reasons: Themes from the Philosophy*

individuals stand in a relation of mutual recognition with others just when they act on principles that others could not reasonably reject.⁶² Thus I can stand in a relation of mutual recognition with you whether or not you in fact recognize that I act on principles that you could not reasonably reject, and whether or not you relate to me in turn by acting on principles that I could not reasonably reject.⁶³ As I am using the term, by contrast, *two* (or more) agents stand in a relation of mutual recognition when there is a *common understanding* (grounded in a common belief) that they each recognise certain normative standards as governing appropriate behaviour between them. This is why I have chosen to use the same label as Scanlon. As I understand Scanlon's view there is nothing necessarily *mutual* about mutual recognition. Yet, as I have been suggesting, I believe that there is something important about a genuinely mutual form of recognition between persons, an idea that is not well captured by the use of other terms.

In light of this, it would be useful, I think, to refer to Scanlon's notion explicitly as *hypothetical* or *notional mutual recognition*. Importantly, that these distinct kinds of relation between people are both valuable is, it would seem, perfectly consistent. Nevertheless, while I believe Scanlon still needs his own notion of (hypothetical) mutual recognition, I also suspect that the appeal of Scanlon's position partly derives from the value we recognise in *concrete* relations of mutual recognition such as those I am describing. For instance, it seems plausible to think that the value of standing "in unity with our fellow creatures" discussed by Scanlon⁶⁴ must extend beyond the hypothetical unity we might acquire by individually acting on principles that we believe cannot be reasonably rejected. Plausibly, it will also require our actually standing in social relations with others in which we jointly recognise the validity of some such principles.

Let me now turn to explicitly consider how the account of mutual recognition I have developed helps us to make sense of the analogy between friendship and the "moral relationship". The central idea is straightforward. We can say that P1 and P2 stand in a relationship of MR_{MORAL} when they share a common belief that they each respond appropriately to the moral reasons they have with regard to each other. For example, P1 and P2 stand in a relationship of MR_{MORAL} when they share a common belief that they respond appropriately to moral reasons not to harm one another, or when they share a common belief that they respond appropriately to certain moral rights and duties they each have.⁶⁵ Notably, on this view, the valuable form of moral relationship in question is only

of Jonathan Dancy, eds. David Bakhurst, Brad Hooker, and Margaret Olivia Little (Oxford: Oxford University Press, 2013), p. 161.

⁶² *What We Owe To Each Other*, p. 162.

⁶³ It is unclear to me whether Scanlon holds that we stand in a relation of mutual recognition with *every other* rational agent when we act on *any* principle that could not be reasonably rejected (as opposed to, say the agents who we could effect by acting on this occasion). I am also unsure whether Scanlon holds that in order to stand in this relation we must act on principles of which it is *true* (or at least reasonably believed) that the principles in question could not be reasonably rejected, or whether it is sufficient that we sincerely believe that we are acting on such principles.

⁶⁴ *What We Owe To Each Other*, p. 163.

⁶⁵ Again, these rights and duties need not be perfectly symmetrical. For example, P1 may owe P2 a promissory duty without P2 owing P1 any reciprocal duty.

realized in our concrete interactions with others. That is to say, we do not stand in a valuable relationship of MR_{MORAL} just in virtue of (say) our shared status as moral agents. Rather, we stand in this relation only when we interact in some way and hold the common belief that we relate to one another appropriately in light of the moral reasons that apply to us.

Someone might object that relations of MR_{MORAL} are not genuine *relationships*. That is because instances of MR_{MORAL} may be one offs – between a customer and a shop assistant, say, or one stranger asking another for directions – and so lack the historical and ongoing components we might think necessary conditions for relationships.⁶⁶ On this view, the mere fact that certain (directed) reasons and duties apply to P1 and P2 with regard to one another does not suffice to place them in a relationship with one another. This objection seems to me to be correct as a matter of our ordinary understanding of relationships. Thus, when we say that two individuals stand in a relationship of MR_{MORAL} in cases where there is no historical or ongoing relationship between them, we must understand ourselves as employing the notion of a relationship in a much broader, and somewhat artificial, manner. However, this does not mean that their interaction does not possess the valuable feature in question. In any case where two agents realize MR_{MORAL} this endows the interaction or relationship in question with a distinct form of value.

More importantly, this account of mutual recognition still enables us to make better sense of the idea that valuable moral relations between individuals in some way resemble valuable friendships. As we saw above, the value of friendship is very plausibly partly constituted by the fact that friends mutually recognize that they each deliberate in accordance with the normative requirements of friendship. Without mutual recognition, a friendship would seem to be a sub-optimal or non-ideal instance of friendship, and perhaps not even a friendship at all.⁶⁷ If it is correct that mutual recognition is central to friendship in this way, it makes sense that we should describe other instances of valuable mutual recognition, including MR_{MORAL} , by appeal to the example of friendship.

Moreover, it seems clear that we are more likely to realize MR_{MORAL} with one another when we stand in ongoing patterns of interaction and relationship – whether because of regular trips to the same hairdresser or doctor, in the context of professional relationships, or indeed, within special relationships like friendship. As these latter possibilities suggest, MR_{MORAL} may be embedded within all kinds of valuable special relationship. Indeed, on some views, our moral duties to one another may be thought of as providing a basis for these thicker forms of relationship, and thus, MR_{MORAL} may be a foundation upon which other kinds of valuable relationship are built.⁶⁸

To summarize, in this paper I have outlined an account of mutual recognition between persons. According to this account two individuals stand in this relationship only when they realize a concrete form of inter-subjective recognition. Specifically, P1 and P2

⁶⁶ Niko Kolodny, “Love as Valuing a Relationship,” *The Philosophical Review* 112 (2003), p. 148.

⁶⁷ At least if there is no mutual recognition between “friends” for some extended period of time.

⁶⁸ See, e.g., Scanlon, *What We Owe To Each Other*, p. 164.

realize a relationship of mutual recognition where they share a true common belief that they deliberate in accordance with the reasons and duties that apply to them in the context of their relationship. I claimed that mutual recognition generally endows our relationships with a distinct form of value, although that this may depend on the underlying value of such relationships. Moreover, I suggested that this account enables us to make sense of the analogy between friendship and the idea of a moral relationship.

If this account of the value of mutual recognition is on the right track, I believe it has important implications for a number of debates in moral and political philosophy. For example, given the requirement that agents have true common beliefs about how they relate to one another, if we wish to support mutual recognition between individuals, we will need to establish social mechanisms and institutions that can serve as a basis for justified beliefs of this kind. This requirement may partly contribute to the justification of formal political and legal institutions, insofar as these institutions instantiate shared normative frameworks upon which to form such beliefs. Furthermore, I suspect that the value of mutual recognition also underpins important facts about the normative structure of our interpersonal relationships even in the absence of formal social institutions. However, an exploration of these issues will have to wait for another occasion.

Suffering, Moral Luck, and the Duty to Self-Care

Alycia LaGuardia-LoBianco

Alycia LaGuardia-LoBianco is a Ph.D. candidate at the University of Connecticut finishing her dissertation on suffering and moral agency.

We owe things to those who are suffering. We ought to help them, seek justice on their behalf, repair any wrongs that have caused their suffering, and so on. That suffering gives rise to moral duties is uncontroversial. Yet the focus on what others ought to do *for* sufferers neglects an equally pressing moral question: what do sufferers owe to themselves?

This question is complicated by the fact that many cases of suffering result from moral luck. Certain contingent causes of suffering – living as a woman in a patriarchal society, growing up with an abusive caretaker, being disposed toward mental illness, experiencing a trauma – are outside of one’s control yet may have a great influence on who one becomes. And this has a moral impact on the duties incurred and the agency possessed to fulfill these duties: bad moral luck may shape, rather than invalidate, one’s moral agency and responsibility.

In this paper I argue that individuals who have been shaped to suffer have an obligation to care for themselves by protecting and expanding their agency. I first describe cases of *complicit suffering*, in which contingent factors shape one such that they behave in ways that contribute to their suffering. I argue that the morally relevant features of these cases are burdened agency and its impact on self-directed actions: one’s ability to help oneself is compromised by the very conditions that generate the need for help. Following the forward-looking perspectives of agency and responsibility advocated by Claudia Card, Lisa Tessman, and Susan Wendell, I hold that victims of bad moral luck are not, thereby, excused from responsibility.

Next, I discuss the motivations for framing cases of complicit suffering in terms of self-regarding duties. One may worry that duties of self-care place an undue burden on sufferers and evoke victim blaming. I respond by arguing that this framing can be a beneficial way for sufferers to approach their situation. I then focus on the duties associated with psychological oppression to frame self-regarding duties. Carol Hay argues that oppressed agents incur the duty to resist oppression for self-regarding reasons. I argue that her views provide a promising start to thinking about duties of self-care in cases of complicit suffering but that these duties must be more finely tuned to the particular difficulties a complicit sufferer faces.

Finally, I put forward my view of the duty to self-care in these cases. Since complicit suffering is marked by damaged agency, sufferers ought to protect and expand that agency in order to prevent further damage and ensure continued self-care. These duties aim not only to block further destruction of agency, but also to empower the agent so they are better able to care for themselves, despite the effects of moral luck.

1. Complicit Suffering

Among the things that are outside of one's control are circumstances that cause suffering. These contingencies are not merely unfortunate; they have moral consequences. Tessman defines moral luck as 'that which is not within an agent's own control and yet affects that agent in a morally relevant way', as when such circumstances impose difficult moral decisions.⁶⁹ But moral luck can also impact *who one is*. This is constitutive moral luck, or luck 'which affects the formation of character'.⁷⁰ In turn, character can influence how and the extent to which someone suffers. For instance, whether one is abused is outside of their control. Yet, abuse has long-lasting effects that can contribute to a character marked by self-deprecating beliefs, feelings of worthlessness, and self-destructive behaviors.⁷¹

Suffering can result from behaviors like rumination, low self-esteem, beliefs of inferiority, self-deprecation, or self-sabotage, that are themselves inculcated by contingent influences, such as upbringing, society, other agents, or traumatic experiences. Call such cases *complicit suffering*, in which an individual's suffering is proximately caused by their behaviors and ultimately caused by forces that shape these behaviors.⁷² The luck at issue here is that of being put in a position to internalize harms that burden agency. But burdened agency does not rescind all responsibility. One does not lack *any* role in who they are shaped to be, and can still be responsible for who they have been shaped to become. Card maintains that moral luck is compatible with both agency and moral responsibility. She distinguishes two orientations of responsibility: a backwards-looking orientation, which attributes blame or praise for past actions, and a forward-looking orientation of taking responsibility for what has not yet happened.⁷³ Moral luck does not dismantle responsibility when viewed from the forward-looking orientation since 'when we take responsibility for something, there is no assumption that we produced it'.⁷⁴

⁶⁹ Lisa Tessman, *Burdened Virtues: Virtue Ethics for Liberatory Struggles* (New York: Oxford University Press, 2005). 13.

⁷⁰ Op. cit. note 1, 14.

⁷¹ This is not to say that behaviors like these are only *ever* caused by external forces like abuse.

⁷² A relevant issue is the degree of moral responsibility and blameworthiness shared by the sufferer and the agents that influenced her, such as her parents or oppressors, and the degree to which victim-blaming may be appropriate. I leave these issues to the side here, but for related discussions of victim-blaming and responsibility, see Susan Wendell, 'Oppression and Victimization; Choice and Responsibility', *Hypatia* 5 (1999), 15—46., Anita M. Superson, 'Right-Wing Women: Causes, Choices, and Blaming the Victim', *Journal of Social Philosophy* 24 (1993), 40—61., and Chapter 3 of Carol Hay *Kantianism, Liberalism, and Feminism: Resisting Oppression* (New York: Palgrave Macmillan, 2013).

⁷³ Op.cit. note 4, 25—27.

⁷⁴ Op. cit. note 4, 29. Tessman complicates Card's view by arguing that moral damage to one's character can sometimes be 'irreversible' (29), and if it cannot be changed, one has to reckon with acting ethically as a 'conflicted self' (22). This conflict can be severe enough to complicate or even prevent one's taking responsibility for oneself (or developing integrity, as Card describes it) since one may not 'fully stand behind such a self' (20). Op. cit. note 1.

Wendell analyzes the responsibility of women living under patriarchy with what she calls the perspective of the responsible actor.⁷⁵ Like Card's, this perspective is future-directed, focused on both an agent's ability to act and their limitations. Even if options are few, Wendell maintains that '[s]ituations are rare in which a person has no choice'.⁷⁶ So though one's character may be intractable, victims of constitutive moral luck can take responsibility for themselves by considering the choices open to them.

The same applies when moral luck shapes one to suffer. Complicit sufferers are impaired in their ability to help themselves, yet may still be responsible for themselves. And this creates a moral difficulty: because they suffer, they owe certain things to themselves. But because this suffering involves damage to moral agency, their abilities to fulfill these duties are impaired. They may either believe they cannot do anything to change their situation or expect themselves to completely alter their characters. Being shaped to suffer thus has serious moral consequences.

In order to tease out these consequences, consider two ways in which complicit suffering damages moral agency and self-care. First, sufferers may have difficulty correctly identifying their own causal roles and the causal roles of others in their suffering. Because self-destructive behaviors are deeply ingrained in the sufferer by formative or traumatic experiences, their ultimate causes are obscured. Additionally, it can be difficult to recognize one's own self-destructive behaviors, especially when focusing on abuses or injustices one has suffered. This tangle of causes makes effective self-care uncertain, obscuring sufferers' obligations to themselves and the obligations of others.

This causal indeterminacy can lead sufferers into one of two extremes: either they fail to see the forces that have shaped them to suffer, and take on full responsibility for their suffering, or see *only* these forces, and fail to see their own role in their suffering. Consider the implications of the first extreme. Sufferers may believe they suffer because of an inherent flaw or failure instead of recognizing the influences that create this belief. In the cases of oppression, Sandra Bartky calls the mistaken attribution of cause mystification, 'the systematic obscuring of both the reality and agencies of psychological oppression'.⁷⁷ Victims are not only made to believe themselves inferior, they are also deceived into thinking this belief is justified.

Sufferers who do not realize how they have been influenced to suffer may also believe their suffering is fixed and change is hopeless. They may not attempt any self-care—why bother trying to fix what is inherently broken? Alternatively, if they feel fully responsible for their suffering, sufferers may have unreasonably high expectations of what they can do to help themselves. They may expect themselves to completely erase their self-destructive behaviors, supposing they are the sole cause and remedy.

⁷⁵ Susan Wendell, 'Oppression and Victimization; Choice and Responsibility', *Hypatia* 5 (1999), 15—46.

⁷⁶ Op. cit. note 9.

⁷⁷ Sandra Lee Barky, *Femininity and Domination: Studies in the Phenomenology of Oppression*. (London: Routledge, 1990). 23.

Failure to acknowledge one's causal role in complicit suffering carries its own problems for self-care. Here, sufferers view other agents as the sole cause of their suffering, so may believe these others are responsible for redressing that suffering. They may *avoid* self-care, refusing to help themselves until others take responsibility for their plight. Alternately, they may feel deterministically shaped by the injustice they have faced, fated to suffer by the forces that influence them and helpless to alter their course.

The second factor that complicates self-care is the emotional response to this causal indeterminacy: sufferers can react with self-blame or self-pity. If they believe their suffering is the result of inherent inferiority, sufferers may blame themselves for their situation instead of attempting self-care. This can reinforce self-deprecating behaviors if they berate themselves for being the way they are – the 'sort of person' who is irrevocably 'broken.' This attitude can reinforce self-loathing and the belief that care is undeserved. Sufferers can take up what Wendell calls the perspective of the oppressor, which embodies a backwards-looking assignment of blame to the victim.⁷⁸ In so doing, one comes to blame and punish themselves for their suffering.

Alternatively, if sufferers do not recognize their causal role in their suffering, they may pity themselves. This can compromise self-care if they take on the perspective of the victim.⁷⁹ This entails placing the blame for their suffering entirely on their abusers. While this perspective frees them from excessive guilt, it can also stagnate efforts of self-care. By focusing on how they have been unfairly made to suffer, they may come to view themselves as powerless rather than as agents. Sufferers who view themselves as victims focus on how they are treated by others, so may not consider what they can do for themselves.

Emotional stability does not guarantee effective self-care: someone who is psychologically healthy may not know what is best for themselves. Self-care can be further compromised by emotional strain, psychological exhaustion, and lack of motivation. Complicit suffering aggravates these factors with a complicated tangle of causes. Given these difficulties, what ought sufferers do for themselves—if anything?

2. Motivations for Duties of Self-Care

Before continuing, I wish to address a potential worry about framing complicit suffering in terms of duties to self.⁸⁰ Given the difficult state that complicit sufferers are in, why think duties of self-care are an appropriate way *for sufferers* to approach their situation? It may seem such duties place an extra moral burden on already burdened agents. And if so, it is not clear that they are beneficial.⁸¹

⁷⁸ Op. cit. note 9.

⁷⁹ Op. cit. note 9.

⁸⁰ I thank two anonymous referees for encouraging me to explain this point.

⁸¹ A third interpretation of the question is: why or how do duties to self arise generally? This issue is unfortunately outside the scope of this paper.

These worries are understandable as they recognize the burdens these sufferers already face without victim-blaming. But rejecting a self-regarding duties framework on these grounds propagates a harmful assumption while also overlooking a potential resource for sufferers. Since the focus is shifted to the duties of other agents, sufferers are overlooked, effectively excluding them from the conversation about moral agency. The implication is that sufferers need help but cannot help themselves, casting a portrait of passive victims too damaged to help themselves and fully reliant on others. The assumption that duties of self-care do not apply to sufferers is therefore *agency denying*, which inadvertently propagates a harmful assumption of passive victimhood.

By contrast, framing complicit suffering in terms of self-regarding duties makes sufferers part of the moral conversation without reducing them to victims: it is agency affirming for sufferers to consider their duties to self-care. Complicit sufferers' damaged views of their responsibility contribute to their suffering. So reframing their moral position as one in which they have duties to themselves by virtue of their suffering is a way to challenge this views. Duties of self-care thus provide sufferers an agent-centered moral perspective that is absent in frameworks focused exclusively on the duties of other agents.

Specifically, there are a few ways in which this framework is beneficial for sufferers. First, because their agencies and views of their own agencies are each damaged, recognizing the agential capabilities they have while also acknowledging their limitations can be agency-building. Recognizing this realistic sense of agency can combat the extremes of feeling utterly powerless and feeling responsible for things outside their control. This thought shares some themes with existentialist ethics, especially as it applies to individuals facing systemically restricted options.⁸² For those who face severe limitations, such as women living under patriarchy or persons of color, to conclude that these situations leave them *no* room to exercise agency is inaccurate, a form of bad faith that attempts to evade the reality of agency. This reality is that there are choices to be made, however small their range, such as choices of how to interpret one's oppression or the commitments one makes under these conditions. An inability to radically change one's human condition thus does not entail a lack of freedom. A framework of duties to self-care serves this end as it casts the sufferer as an agent while remaining sensitive to the burdens that restrict agency. On this framework, burdens are part of agency, not preclusions to it.

Second, a framework of duties to self provides a more objective way for the sufferer to evaluate themselves and their situation. This can help correct their distorted views of self-pity and self-blame. Through a framework of duties, a sufferer may be able to discern what they are realistically responsible for and capable of changing. A moral evaluation of the situation will yield a more objective verdict since it is framed by standards of right, wrong, and duties, rather than by feelings of pity, blame, and defeat. And by seeing the nuanced moral reality of their suffering, rather than a self-defeating distortion, they can respond to it in a proactive way.

⁸² See, for instance, Beauvoir 1948.

A complicit sufferer may believe that they deserve the suffering they bear. This is why self-harming behaviors are so easily reinforced: part of what a sufferer has internalized is the belief that they are a failure or helpless, and these beliefs beget confirmation that they are worthless, wretched, and ultimately undeserving. And this is exactly the view that duties of self-care challenges. If the *moral* response to their situations is a duty to help themselves, this means sufferers deserve and owe themselves care.⁸³ This is not a prudential consideration to maximize happiness, but a moral response demanded by their situation. Framing their suffering in these terms can help sufferers recognize that, contrary to their beliefs, they deserve better. Complicit sufferers are *just* the people who need this guidance because of how they suffer.

The self-damaging beliefs and behavior characteristic of complicit suffering suggest that sufferers can benefit from a framework that challenges these beliefs, affirms their agency, and counters the damage. These agent-centered reasons ground the duties of self-care in cases of complicit suffering, both motivating the framework and justifying the duties themselves.

3. Self-Regarding Duties

Sufferers' duties of self-care are part of a larger issue of duties of compromised agents. A difficulty lies in determining whether such duties arise given that agency is compromised, and if they do, how well these duties can accommodate the compromises. Before giving my own view – that though sufferers' agencies are damaged, they have duties of self-care tailored to their burdens – I will look at one other way in which duties of self-care in compromised agents have been addressed: Hay's account of self-regarding duties to resist oppression. Doing so will help show what is required in such an account and provide a starting point for modeling these duties.

Hay's account of the self-regarding duty to resist oppression offers a way to understand moral agency when rational nature is damaged by oppression.⁸⁴ According to Hay, a Kantian framework of duties to self, and in particular the duty of self-respect, is uniquely positioned to explain the oppressed individual's obligations to protect themselves against harms of sexist oppression.⁸⁵ Because the duty of self-respect 'is an obligation to recognize the value of the rational nature within us and to respond accordingly', and since oppression threatens just this rational nature (for instance, by undermining the ability to set and meet one's own ends), the oppressed is obligated to protect that rational nature against these harms by resisting.⁸⁶

Since resisting oppression at every opportunity may be impossible or exhausting, Hay argues the duty to do so is imperfect: it can be discharged by various actions. This

⁸³ I thank [name removed for review] for this suggestion.

⁸⁴ Carol Hay, 'The Obligation to Resist Oppression', *Journal of Social Philosophy* **42** (2011), 21—45.

⁸⁵ Carol Hay, *Kantianism, Liberalism, and Feminism: Resisting Oppression* (New York: Palgrave Macmillan, 2013).72-78.

⁸⁶ Op. cit. note 35.

permits latitude in which actions to take. One can resist externally, say by opting out of oppressive norms or institutions or participating in activism, but resistance can also be internal. An oppressed person can become ‘the sort of person whose rational nature was simply not damaged by oppression, perhaps by building up mental walls against many of the harms’ to fulfill their duty.⁸⁷ Internal resistance is limited, however. Hay notes that internal resistance will not always be sufficient for resisting since ‘[m]ost people’s psychologies are simply not oppression-proof’.⁸⁸ Additionally, the duty to resist permits latitude in refraining from action. Hay argues that damage to rational nature from oppression is like the gradual erosion of a rock. While one drop of water will not make a noticeable difference, a flood of continuous drops will. Likewise, one oppressive incident may not be enough to dismantle rational nature, but constant unprotected exposure will. So, the duty to resist need not be constantly upheld, so long as one does not *always* leave their rational nature unprotected.

However, Hay’s characterization of internal resistance as building ‘mental walls’ against oppression stays silent on what resistance with weakened resources would be like.⁸⁹ How does someone build up these walls among the psychological detritus of inferiority, low self-confidence, or self-loathing? How must this construction work differently given that one is not the ideal builder? I agree with Hay that such internal resistance is possible, but it will not be as simple as ‘becoming someone who is not harmed by oppression’ if one *has already been harmed by oppression*.

Think again of Hay’s erosion analogy: a single harm of oppression will not damage one’s rational nature, so protection of rational nature requires that one does not allow harms to accumulate. This means the primary concern is protecting rational nature from *further* damage due to oppression. But, in cases of complicit suffering, one’s rational nature has already been significantly damaged by internalized harms. Sufferers are in the position of protecting their already damaged rationality from the harms *they* perpetuate. This is a significant difference: fortifying a defensive wall will be much harder if one is already under attack from within.

Finally, since Hay’s account allows that the duty to resist oppression can be discharged by many different actions, this means that when a particular act of resistance is too burdensome, agents can discharge their duty through less taxing actions of internal resistance. While this accomplishes Hay’s goal of avoiding onerous duties, the effort to mitigate burdens depicts victims as diminished moral agent: certain actions are no longer obligatory when they are too taxing. At that point, agents are still obligated to resist, but only to the extent that they are capable. So, while the duty to resist is always in place despite burdened agency, the actions that discharge it *scale* with the degree of this burden.

⁸⁷ Op. cit. note 36.

⁸⁸ Op. cit. note 36, 144.

⁸⁹ Op. cit. note 36, 141. Hay does note that one can avoid internal resistance because of self-deception, so she does not think that internal resistance is impervious to the damage wrought by oppression. But her discussion focuses on what counts as a legitimate avoidance of internal resistance, not the building of this resistance.

While Hay's 'scaling' approach is careful to accommodate the limitations of the agent without rescinding their duty, it does so at the price of recognizing them as *full* agents. Rather, less is morally required of them because of their burdens. Additionally, since less is required of them, the moral demands of their situation go overlooked. If the actions that discharge their duty attenuate when they become too difficult, this implies that their moral situation requires less of them. But their situation has not changed – their rational nature is still under attack, and their duty to protect it just as imperative. Why, then, should they only be required to discharge this duty with the easiest action available?⁹⁰ Hay's account can be bolstered with action-guidance for extremely burdened agents.

4. Duties of Self-Care

Because complicit suffering is defined by burdened agency, *this situation morally demands that sufferers protect and expand their agency*. Even though agency is impaired, this does not mean sufferers have no agential resources. Confidence, resilience, time, energy, relationships, support systems, willpower, courage, temperance, self-restraint, expressing or regulating emotions, willingness to reach out to others, motivation, and introspection are all exercises of agency that can be used to protect and expand the impaired components of agency. With the agential tools they already possess, sufferers can discharge their duty of self-care.

When agency is most burdened, driven by guilt, pity, and feelings of inferiority, a sufferer's duty is to stop the onslaught of damage. Given that their agential guard is already down, they ought to prevent further destruction of their agency: they should use whatever resources they have to prevent their situation from becoming worse. This is akin to performing damage control, harboring the resources they have and protecting them from further depletion. And even in their compromised situation, they can discharge this duty by a few different tacks.

First, they ought not further deplete their agential resources by self-sabotaging. By avoiding situations and behaviors that would place a further burden on their already strained resources – surrounding themselves with people who deprecate them, alienating others who could help them, and so on – they can help prevent their resources from depleting further. In addition, they should refrain from directly harming themselves. This includes avoiding psychological harms like being cruel to oneself in some of the ways one can be cruel to another: calling oneself 'worthless' and 'unlovable'. These sorts of harms undermine their target's self-esteem and emotional stability by design.

Another way to protect their agency is to reach out to others.⁹¹ Drawing on the care of others can help fill in the gaps of impaired self-care, whether this involves

⁹⁰ I thank an anonymous referee for pushing me on this objection.

⁹¹ I thank a referee and the audiences at [name of conferences removed for review] for encouraging me to develop this point.

recruiting others to help them avoid self-sabotage, to provide a healthy distraction, to give them an objective perspective on their situation, or to remind them of their abilities when they feel most incapable. Again, this is in the service of protecting their agential resources from further depletion when they are already impaired, so while others can provide general support, they can also help in this capacity, especially if sufferers lack confidence in their abilities to protect their agencies while in crisis.

If sufferers are not in the midst of a crisis, yet still struggling with impaired agency, their pressing duty is to expand their agency. This duty includes recognizing the extent and limitations of their agency. In other words: they ought to endeavor to become aware of their agential impairments and resources in order to build on their strengths and reinforce their weak spots. This may begin by considering that, perhaps, there are reasons greater than themselves that explain their suffering, or, perhaps, there *are* some ways that they act self-destructively. If one realizes that they were mistaken about the cause of their suffering, this may inform the best response to it.

Recognizing the extent of one's agency will also involve untangling the distorted causal knots of complicit suffering: they must identify the boundaries of their agency as well as they can if they are to fortify it. Others, especially other complicit sufferers, can be a great resource in this regard. Recall that sufferers can easily fall into the extremes of self-blame or self-pity, both of which can hinder agency. Gaining a perspective external to themselves can help correct these distorted views, especially when this perspective belongs to someone who has gone through a similar struggle. This may count as evidence that they are not an aberration: others have been shaped to make themselves suffer, too, and have nonetheless helped themselves. And this may show both that they are not entirely at fault for their past, yet may have more control than they realize over their future.

Further, others can help motivate the project of self-care in the first place. This is especially needed if sufferers are not moved by the desire to help themselves or do not believe they deserve care. Even if others cannot inspire intrinsic motivation for self-care, they can help sufferers go through the motions until they believe they deserve it, hold them accountable for self-care, or give them an external reason to make an effort at self-care. Other sufferers are resources in this regard, as they can help sufferers recognize the roots of their feelings of worthlessness while demonstrating that if someone else deserves care in this case, and if it is structurally similar to their own case, then they, too, deserve care.

Another component of expanding agency is learning to recognize the effects of certain behaviors on oneself, as this can be a clue in learning where one's agential powers and deficiencies lie. This may also involve educating oneself on one's own behaviors, starting by noticing one's behaviors and what results from them. If one recognizes the mood that results from dwelling on their guilt and that which results from intervening on a spiral of self-blame, say, one can come to appreciate how certain resources can foster agency. From this awareness of the resources *as* resources, they can find ways to protect and expand them. Even if lacking full awareness of her behaviors, sufferers can endeavor

to learn more. This may be more easily accomplished with the help of others who can point out these behaviors.

Finally, the duty to expand agency involves preventative measures in order to ensure continued protection of agency. This is akin to building an agency emergency plan, a base of reliable resources that one can deliberately draw on when all other options are exhausted.⁹² The tactic is to have some ‘fail-safe’ in place in the case that it is needed. Additionally, preventative measures can include setting up the conditions that ensure effective self-care or anticipate weaknesses when emotionally taxed, say by asking a friend to step in should they exhibit certain behavior. A final component of expanding agency is the regular exercise of certain resources, such as reminding oneself that they are not a pure victim or are worthy of compassion, so that these resources are maintained and are ready when needed.

5. Conclusion

The existence of suffering gives rise to questions of ethical action guidance. However, the other-regarding focus of many these questions overlook the moral agency that sufferers retain. By viewing sufferers as moral actors rather than victims to be acted upon, we can expand our moral action guidance to include burdened agents.

⁹² One instance of a crisis plan is the Wellness Recovery Action Plan. Mental Health Recovery. 1995-2016. WRAP is. <http://mentalhealthrecovery.com/wrap-is/> (accessed May 3, 2017).

Grace and Alienation

Vida Yao

Love is knowledge of the individual.

– Iris Murdoch⁹³

I. INTRODUCTION

I will here respond to one puzzle that emerges from thinking about how to love an individual person well. It is a familiar thought that loving another person well involves seeing him in a positive light. At the same time, according to an attractive understanding of love, to love another person well is to *attend* to him, where the goal of such attention is to *see him as he is* – that is, to see him as he *truly* is. And so, here is the puzzle: are we to see our loved ones as better than they are, or are we to see them as they truly are?

Some who have emphasized the importance of *attentive* love might deny that this question is genuinely open. Given close and generous attention to another person, we will not only achieve a more truthful view of him, we will also find the value or goodness within him that ultimately grounds and justifies our love.⁹⁴ This is an attractive moral ideal. The worry, put simply, is that with certain people it may not be true.

If it *isn't* true, how then, are we to love those particular people well? The solution I propose here will involve articulating one way to understand an aspect of interpersonal love that – though central to Christian faith – can be exercised and felt toward one another in strictly human circumstances, and regardless of the truth of that worldview. That aspect of love is love's *grace*.

I will here propose a secularized conception of grace, according to which grace is a love of the qualities of *human nature*, even when that nature is uncorrected by virtue. Given that, the qualities that grace is responsive to will include qualities of our natures and our characters that *aren't* good qualities, but which are nonetheless *human*. Given that it is a love of a person's qualities, it is importantly different – in a way that I will elaborate on – from love understood as justified by a person's bare humanity, or by the relationship that she stands in with you.

⁹³ “The Idea of Perfection”, *Existentialism and the Mystics*, 321.

⁹⁴ As Iris Murdoch expresses this idea: “If apprehension of good is apprehension of the individual and the real, then good partakes of the infinite elusive character of reality”. “The Idea of Perfection”, 333.

I will first argue that this form of love is *desirable*, by showing how it is uniquely suited to certain difficulties we face in loving another person well – precisely in those cases where closer attention to the beloved does *not* reveal his good qualities, but instead brings his bad ones into starker relief. I will then provide a positive conception of grace that can meet these difficulties.

I.I The Tension

Consider the proposal that one centrally important and desirable aspect of love, discussed by Iris Murdoch and elaborated on by David Velleman and Susan Wolf, is that it is *attentive*, where the goal of this directed attention is to see the beloved *as he truly is*.⁹⁵ According to Murdoch, loving another person well involves moral progress, involving an exercise in our ability to see beyond ourselves and really *attend* to reality. The bad qualities that we see in one another result from our failure to do this: rather than seeing a person as he is, we see him tainted by our fears, prejudices, and fantasies. But when we engage in the moral and perceptual improvement necessary for seeing how things really are, we remove this distortion and come to see something *good* about the object of our attention.⁹⁶

While a morally attractive thought, we might ask whether we are entitled to the harmony that this picture proposes. In attending to a person as he *really* is, why think that we are guaranteed to find anything *good*? *Prima facie* it seems that close attention, in revealing reality to us, can thereby reveal to us the *bad* parts of reality. Imagine, for example, looking more carefully at a person, and seeing the political and cynical motives underlying friendliness and charm.⁹⁷ A tension, then, arises between these three claims:

- 1) Our love, if a rational response to the qualities of our beloved's self, is a love for *only* the good qualities of our beloved's self.

⁹⁵ (Wolf 2015)

⁹⁶ Her case of M and D illustrates this idea: M comes to see D as she *truly is*, which is refreshingly simply and not obnoxiously vulgar. But in order to do this (at least on one reading) M must overcome e.g. *her* (classist) prejudices. As Wolf notes, there is more ambiguity in Murdoch's meta-ethics than I have let on here. Nonetheless, I will treat Murdoch as endorsing the idea that perception of the reality of a person will guarantee perception of something good about that person.

⁹⁷ Given Velleman's Kantian framework, it is just a matter of conceptual truth that a person will have something valuable – indeed immeasurably valuable – about him. It is less clear to me exactly how goodness is guaranteed by Murdoch (or whether she really endorses the guarantee). On Carla Bagnoli's reading of her, Murdoch is not a standard realist about value, (nor anti-realist!) as goodness for Murdoch involves the use of our representational capacities, but also involves seeing value in the world as something achieved *through the use of* our representational capacities. And so, upon the proper use of loving attention, I should be able to *achieve* a vision of value which will also be an accurate one. I do not think one needs to be a Platonist like Moore to generate a tension between seeing accurately and seeing goodness, but I will not try to provide a positive realist view here. (Though perhaps one example of someone who holds such a view is Susan Wolf). Wolf, "Good-for-nothings."

- 2) There is no self that is wholly good; characters are a mix of good, bad, and neutral qualities.
- 3) There are circumstances in which love must be *fully attentive*, in that it is felt given extensive knowledge of the qualities of the beloved's self, and be love for more than the good qualities of the beloved's self.

I assume that (2) needs no defense. After elaborating on and defending (3), I will then propose a conception of grace that can meet the need expressed in (3). This conception will also give us reason to reject or attenuate our commitment to (1). One might accept (1) either because reasons to act or feel positively about something, in general, are responses to *goodness*, or because one is attracted to a view of love where it is not a response to the *goodness* of our beloveds, because not a rational response, at all. One goal in this paper is to show that grace bears the marks of a rationally-evaluable attitude, in being responsive to the qualities in the beloved – though it is responsive to more than goodness. But before getting there, it will first be necessary to motivate the proposal: to see why grace, understood *both* as a rational response to the qualities of the beloved, though not limited to the good qualities of our beloved is desirable, by first showing why (3) is true.

I.II The Need for Fully Attentive Love

Though there are good forms of love that do not demand full attention to the beloved's self there are those who may need fully attentive love, for good reasons. But what exactly do I mean by love being "fully attentive"? And what exactly is desirable about fully attentive love? By fully attentive, I mean to emphasize two things about this kind of love. The first is that it is affection felt in extensive knowledge of the qualities of a person's self, or character, or personality. The second is that the qualities that one loves about the beloved's self are also extensive: it is not just the beloved's *good* qualities that one loves.

To illustrate the sort of context in which fully attentive love may be desirable or necessary to give and receive, consider this scene from Marilynne Robinson's novel *Home*, in which the main character, Glory, reflects on the quality of her brother's soul while helping him wash up from a night of heavy, self-destructive drinking:

Glory said, "You might rub your hands with shortening. That would probably dissolve the grease..." She took the can from the cupboard, scooped out a spoonful, and put it in his palm. She said, "Remember when you talked to me about your soul, about saving it?"

He shrugged. "I think you may be mistaking me for someone else."

"And I said I liked it the way it is."

"Now I know you're mistaking me for someone else." He did not look up from the massaging of his hands.

“I’ve thought about what I should have said to you then, and I haven’t changed my mind at all... your soul seems fine to me. I don’t know what that means either. Anyway, it’s true.”

He said, “Thanks, chum. But you don’t know me. Well, you know I’m a drunk.”

“And a thief.”

He laughed. “Yes, a drunk and a thief. I’m also a terrible coward. Which is one of the reasons I lie so much.”

She nodded. “I’ve noticed that.”

“No kidding. What else have you noticed?”

“I’m not going to mention vulnerable women.”

“Thanks,” he said. “Very generous in the circumstances.”

She nodded. “I think so.”

He said, “I am unaccountably vain, despite all, and I have a streak of malice that does not limit itself to futile efforts at self-defense.”

“I’ve noticed that, too.”

He nodded. “I guess there’s nothing subtle about it.”⁹⁸

This scene marks an important moment in the novel: an affirmation of Glory’s attitudes toward her brother, Jack, who she has slowly grown to know after a long estrangement. Earlier in their re-acquaintance, Glory had noticed and remarked that she, as she puts it, *likes his soul the way it is*. Given the sort of person Jack is, and the person *Glory* is, she was puzzled by this.⁹⁹ Now, explicitly within a context in which Jack has once again failed to live up to values and standards of virtue they both endorse, Glory finds it important to re-affirm and express her sentiment once again. Though Jack tries to deflect the comment by suggesting that Glory doesn’t really know what his soul is like, the rest of the passage reveals that her affection toward him is paired with a discerning awareness of his vices and the distinctive ways in which they manifest in his behaviour. Notice, too, that rather than

⁹⁸ Marilynne Robinson, *Home* (New York: Farrar, Straus & Geroux, 2008), 287-288.

⁹⁹ Glory is both notably morally good *and* pious. We have also learned by now that one of Jack’s greatest wrongs – fathering and then abandoning a daughter – deeply affected Glory, who grew to love the child before she died of an infection. We also learn that Glory’s dreams of a simple family life had been dashed by a man who, like Jack, took advantage of a “vulnerable woman”.

disproving his claim that she doesn't know him by citing the good qualities one might argue he possesses, she alludes to some of his worst.

Jack serves as a challenge to Glory. Their re-acquaintance begins tenuously, and their relationship grows in starts and fits as Jack reveals and confesses his vices, always with the wariness and expectation that at some point his sister – like others – will either have to direct her attention elsewhere, or retract her love.¹⁰⁰ He eyes her affection and kindness with suspicion; she experiences doubts and questions about how exactly to respond someone who has brought so much strife and sadness to their family, and to others. Nonetheless, it seems as though there must be a way for Glory to love her brother, and love him *well* and *attentively*, even though more familiar forms of love fail to meet the challenge. The desirability or “need” for fully attentive love in this case is one relative to both the possibility of their particular relationship, as well as a need that Jack has in light of his otherwise lonely, alienated, and disconnected existence from other human beings.

One way of understanding the challenge is that there is just a straightforward conceptual difficulty in making sense of loving a person's soul, “the way it is”, while *also* being aware of its bad features. Consider Kate Abramson and Adam Leite's conception of love, as “an affectionate attachment appropriately felt as a non-self-interested response to particular kinds of... features of character expressed by the loved one in interaction with the lover”.¹⁰¹ However, the features of character that Abramson and Leite believe love is an appropriate response to are “morally laudable ones” – and these are precisely the ones that, when it comes to Jack's soul, have tenuous footing.¹⁰²

But there are other sorts of reasons or justifications that we can cite to make sense of why we love the people we do, even when their souls are in a bad state. As Niko Kolodny has argued, a paradigmatic sort of love of this type is the love we feel toward those we have long-standing relationships with, where those relationships have developed over time with those particular people.¹⁰³ One can love a person in this way even if there is nothing about the *quality* of that person's character that one thinks of as *morally good*, or *good* at all – it can be enough that one has shared a history with her.

The problem with this answer is not that this sort of love is that it isn't valuable, or that isn't “real” love; it is a real form of love, and involves an admirable kind of interpersonal

¹⁰⁰ The scene continues, with Jack testing the limits of her affection by elaborating on what he calls his “mortal sins”. He clarifies to Glory that for some time of their estrangement he was in prison, not jail, as he had her previously believe. The scene concludes with her response: “I don't care if you were in prison.”

¹⁰¹ (Abramson and Leite 2011, 677).

¹⁰² One might think that Abrahamson and Leite should expand their conception of love to all of the *good* qualities of a person's character, and not just the *morally good qualities*. As Susan Wolf has argued, we sometimes love people because of their excellent qualities when those qualities are not exactly *moral*, and can come into conflict with morality. Though I think that is plausible, it will still not help to solve Glory's problem with Jack, given what his character is like and given the demands of fully attentive love. The problem with Jack is not that he lacks morally excellent qualities and has other excellent qualities, it is that he has a host of *bad* qualities (both moral and non-moral).

¹⁰³ (Kolodny 2003)

loyalty. The point is rather modest: in certain interpersonal contexts, a person might need more. Someone like Jack will need a kind of love that is attentive and responsive to *who he is*, not just what relationship he stands in or the history one shares with him.¹⁰⁴ He isn't comforted, for example, by the love of his other siblings who do indeed love him because he is their brother. Rather than drawing the siblings together, this kind of love only alienates Jack further – he comes to resent it, alongside the benevolence or obligation it is paired with.

Perhaps Glory could focus her attention on more positive interpretations of Jack's character. As Sarah Stroud argues, there are many ways in which we can come to see another in a more positive light, some of which result from an epistemic partiality that good friends exhibit to one another. So Jack is a thief, but perhaps he only steals out of need. He leaves hurt feelings in his wake, but that is just the result of a bracing honesty.

While there is something valuable in being partial or charitable toward our loved ones in this way, or having faith that they are really good after all, the challenge, again, is that in certain contexts, in certain relationships, something else may be needed – a kind of love that isn't based on partial or charitable interpretations of who one is, but on what one's character is really like.¹⁰⁵ Like with his relationships with his siblings, it is notable that those who have expressed their faith in Jack in the past only seem to alienate him further, and that his father's love (which when affectionate, is based on a positive or charitable interpretations of his son) only leaves him feeling more isolated.

What about the idea that Glory loves Jack because Jack – like all rational beings – has the value of being a *person*? According to David Velleman, *all* persons – even those with the kind of character that Jack has – are worthy of love, simply in virtue of their rational will. One might object to Velleman by *denying* that there is an essential and valuable feature that all rational beings share, or by suggesting that such a capacity can be manifested by *bad* qualities which are not proper qualities for love to be responsive to, or by arguing that when we want to be loved we want to be loved for the qualities that differentiate us from other beings.¹⁰⁶ I want to raise a different worry.

¹⁰⁴ Glory herself considers the idea that her love of Jack is a love of her brother. But she contrasts considers her love of her siblings, and then considers her affection for her “irksomely mature and responsible” sisters, also noting that, “fond as she was of [her two other brothers], it was hard to remember that they had ever seemed marvelous to her” in the way that Jack did, and still does.

¹⁰⁵ Notice, too, that Glory's love isn't based on faith in Jack: as she emphasizes, she likes his soul *the way it is*, not the way that she expects or hopes for it to become, or his “true” soul that she believes is good deep down inside. In contrast, faith is expressed by another character as a way of thinking about Jack: she suggests that God's grace is a matter of understanding that *everyone can change for the better*. Perhaps as a way of answering whether this is what explains Glory's affection for him, Jack immediately go off the rails once again, drinking himself into oblivion. It is *after* this happens again that Glory takes it to be important that she restate her sentiment.

¹⁰⁶ Abramson and Leite raise the point about sadism against Velleman. Benjamin Bagley raises the point about the importance of the irreplaceability of the loved one. As he writes, “the resulting ideal of love is not only demanding, but strangely impersonal. It collapses, or at least trivializes the distinction between love for

For love to be fully attentive in the sense that I am interested in, it must be love for the *qualities* of our selves – and it must also love those qualities in the right way. Velleman does accommodate the sort of qualities I am interested in, but it is important to see what role he thinks they play in love of another person:

The immediate object of love, I would say, is the manifest person, embodied in flesh and blood and accessible to the senses. The manifest person is the one against whom we have emotional defenses, and he must disarm them, if he can, with his manifest qualities. Grasping someone’s personhood intellectually may be enough to make us respect him, but unless we actually *see* a person in the human being confronting us, we won’t be moved to love; and we can see the person only by seeing him in or through his empirical persona. (Velleman, 371)

To couch my point in Velleman’s terminology, the thought is that fully attentive love is love of a person’s “empirical persona” – or at least, those aspects of one’s empirical persona that constitute her *self* or her *character* – but not as a necessary means to loving the personhood that underlies that persona and which is only accessible to us intellectually. One might not want to be loved solely because of the color of her hair (an example that Velleman uses), or her flesh and blood (literally speaking), but this is altogether different from being loved because of one’s sense of humour, love of animals, or taste in music – that is, those qualities that can be attributed to a *personality*.¹⁰⁷ And one might want to be loved for these qualities not because they are expressive of a deeper value that lies underneath these qualities, or because they are the necessary means that a person must love in order to love this deeper value – one wants to be loved for these qualities because they constitute *who we are* – at least in this world.¹⁰⁸

These conceptions of love will fail to meet the kind of challenge that Glory faces. But why should we think that there is anything reasonable about the demand that one’s love be fully attentive, or anything good about the kind of love that is able to satisfy this

particular persons and love for humanity as such” (486). As I will discuss, I think that Velleman’s proposal is actually “impersonal” in at least *two* ways.

¹⁰⁷ Bagley suggests that the qualities of a person’s self that are the right sort of reason to cite in love are those that are revelatory of a person’s values. While I think we are interested in the same kind of quality, I would want to emphasize that these need not be expressive of a person’s *values*, but just of her personality. These qualities can include, for example, a person’s vivacity or vitality, her perspicacity, her decisiveness or mental steadiness, and certain forms of creativity.

¹⁰⁸ In emphasizing the self as the object of love’s attention, I am in sympathy with part of Bernard Williams’ position in “Persons, Character, and Morality”, which resists the idea that the best way for us to relate to one another is to relate to one another in virtue of universal aspects of who we are.¹⁰⁸ I would suggest that this is because universal characteristics (like our bare humanity) are not the entirety of what constitutes a person’s self or personality, and so love of just that bare feature of us will not be fully attentive. Unlike Williams, I am less concerned that the differences in our characters make us *unique*, and so *irreplaceable*. My point is that sometimes we want to be loved for what we are qualitatively like, where these qualities should not be confused with just those qualities of, for example, our bodies, or the roles that we stand in to one another. And we want to be loved for these qualities *regardless of whether or not they make us unique*.

demand? While some may positively desire forms of love that are not fully attentive, those of a more critical and self-reflective bent, and who have only tenuous relationships to goodness, may find themselves needing of something else. Jack falls into this latter group. When others' assessments of him are out of step with a truthful understanding of himself, he experiences an isolating disorientation.¹⁰⁹ As I noted above, he feels uneasy with his father's affection, precisely because he knows that such warm feelings are based on an inaccurate or incomplete understanding of what he is like. Even those who are well-intentioned and have faith in Jack, or who focus on the good aspects of his character, will either fail to see him as he truly is (instead, focusing on an idealization of who he is), or focus their attention to something within him that he cannot trust to be a consistent feature of his character. Given this, he experiences these forms of love as looking *past* who he really is, to an image he has no hope of living up to.¹¹⁰

In contrast, Glory does not skirt around the facts of Jack's past and the conclusions that one would justifiably draw about Jack's character. And the fact that her response is one of affection and warmth, rather than hostility, indifference, moral criticism, or evasion, allows her to speak truthfully about him without the harshness or disappointment that have made the truth so painful to him, and others. But more importantly, because Glory's love is one that is directed toward Jack as he actually is, and does not depend on good qualities that may or may not persist within him, Jack can *trust* her love in a way that he cannot trust the love of others. In trusting, and so letting down his guard with Glory, Jack is given the chance to connect with another person who truly sees him as he is. She shares his painful vision of himself, with affection and tenderness rather than hostility. It is through her love that Jack retains *some* connection with the rest of humanity, who has (perhaps justifiably) written him off.

II: GRACE

But how should we understand the nature of this kind of love? Is Glory's love for Jack, in being responsive to more than his good qualities, and not predicated on charitable interpretations or faith, or on the value of her relationship with him, a "blind" or arational attitude? We might be attracted to this given that one might have the same thoughts about *God's* love toward his created beings, as well: one reason why grace is so mysterious, we might think, is that it's ultimately *arational*. But I think that there is more to say about the

¹⁰⁹ For example, earlier in the novel Jack rejects Glory's attempts to excuse his past behaviour because of his youth: "No I wasn't young. I don't believe I ever was young. Excuses scare me, Glory. They make me feel like I'm losing hold" (154).

¹¹⁰ Preston-Roedder has argued for the importance of faith not only in our loved ones, but in ourselves as well. Perhaps if Jack did have faith in himself, he wouldn't be so alienated by standard forms of love. I agree. But some people simply *don't have this faith* in themselves, and so, grace can be understood as an attitude that *supplements* faith in cases where faiths runs out – either on the part of the beloved, or on the part of the lover.

grace we sometimes feel toward one another that will bring it closer to something like a rationally-evaluable attitude.¹¹¹

In his discussion of grace, Robert Adams begins by noting that grace is, or is an essential aspect of *all* genuine forms of love. This aspect is love's *non-proportionality*. As Adams writes: "In a preliminary formulation, we can say that grace is love that is not completely explained by the excellence of its object. Within certain wide parameters, at any rate, it is not proportioned to the excellent of its object, nor conditioned on the degree of that excellence. To the extent that that degree can be measured, grace typically outruns it" (151).

I agree with Adams that all genuine forms of interpersonal love are non-proportional, and so are all gracious in his sense. But I will use the term "grace" to focus on a particular kind of attitude, focusing on forms of love that are non-proportional, where that lack of proportionality is explained by the fact that this attitude, though responsive to qualities of the beloved, is not exclusively responsive to the *good* qualities of the beloved. This is why this love "outruns" the beloved's excellence.

The proposal is that grace is non-proportional to goodness because it is responsive to qualities of *human nature*. The qualities that make up human nature are *not* always qualities that we do or should think of as *good*. But even though qualities of human nature aren't wholly good, or may even be bad, it is still possible to become endeared to an object upon noticing its possession of such qualities. Consider Angel Clare's first experience of Tess Durbeyfield's beauty:

He had never before seen a woman's lips and teeth which forced upon his mind with such persistent iteration the old Elizabethan simile of roses filled with snow. Perfect, he, as a lover, might have called them off-hand. But no — they were not perfect. And it was the touch of the imperfect upon the would-be perfect that gave the sweetness, *because it was that which gave the humanity*.¹¹²

As another illustration, take into consideration the affection that Rosalind Hursthouse describes in her discussion of "arational" actions. As Hursthouse elaborates, while some

¹¹¹ I want to begin by making clear that it is no part of my goal in providing *one* conception of grace to capture all of the ways in which it is perfectly good to use the term. In striving only to describe one kind of grace, there will be some meanings preserved and emphasized, others that will come to be only accidentally related to the target phenomena, others simple abandoned. Along with being influenced by the *content* of Murdoch's moral philosophy, I present the work here as an exercise in moral philosophy using some of Murdoch's proposed methodology: my task here will be to provide what she calls an "exploration" of grace, rather than an analysis of it.

¹¹² Thomas Hardy, *Tess of the D'Urbervilles*. (New York: Dover, 2001) 119, my emphasis. This is not to say that Angel's attitudes toward Tess (either here, or ever) constitute a genuine or desirable form of love; the example is simply meant to illustrate the way in which *human nature* can itself be attractive and endear us to objects.

aspects of our emotional lives can be “rationalized”, and so made valuable through this process, other aspects simply cannot. Examples include actions explained by anger, hatred, or jealousy, such as: “violently destroying or damaging anything remotely connected with the [hated] person... e.g., her picture, letters or presents from her, awards from her, books or poems about her; the chair she was wont to sit in, locks of her hair, recordings or “our” song, etc.” Hursthouse notes:

We might well find something rather touching or endearing about people's performing many of the arational actions; even the disturbingly violent ones seem to evoke some sort of bond of sympathy. When I have read this paper to discussion groups, I have found that the list of the examples at the beginning always provokes instant delighted recognition; everyone knows what it is like to act in some of these ways, and is somehow pleased to hear it acknowledged and described.¹¹³

The pleasure or delight that Hursthouse notices need not only be explained by a bond of sympathy over the *specific* action or motivation that one is responding to: one need not, for example, only feel this affection in light of her *own* experiences of anger, hatred, or jealousy. The basis of the bond is broader than this – it can be felt toward nearly *any* typical human failing, even if one lacks first-hand experience with it.

In rejecting the idea that the concept of the human being is not the concept of something that is necessarily *good*, I am not thereby committed to the idea that grace is responding to something “purely descriptive” or “non-normative” – for example, qualities of the human species understood as a biological concept.¹¹⁴ I also do not mean to pick out one abstract capacity we might identify as being essential to *personhood*, such as the capacity to reason or set rational ends. So what do I mean to be picking out?

Here I can only provide the contours of further investigation. Consider a point emphasized by Philippa Foot that on an Aristotelian picture, virtues are best understood as *correctives* to tendencies that one finds in man (or human beings). She writes:

...going back to the idea of virtues as correctives, one may say that it is only because fear and the desire for pleasure often operate as temptations that courage and temperance exist as virtues at all. As things are we often want to run away not only where that is the right thing to do but also where we should stand firm; and we want pleasure not only where we should seek pleasure but also where we should not. If human nature had been different there would have been no need of a corrective disposition in either place, as fear and pleasure would have been good

¹¹³ (Hursthouse 1991)

¹¹⁴ In a discussion of ethical concepts and a rejection of a clean and complete “fact/value” divide, Cora Diamond notices the difficulty in claiming that “human being” could be a concept of vital ethical importance; she too emphasizes that she doesn’t mean by this either the concept of a *good* human being or the biological sense of human being. Diamond, “Losing Your Concepts”, *Ethics*, 98:2 (1998).

guides to conduct throughout life.¹¹⁵

So, in investigating virtues and vices, we already (implicitly) have in mind some idea of what human nature is like. And unpacking this idea further provides another way to think about the qualities we will arrive at when we spell out this idea explicitly: when we think about the virtues as correctives to human tendencies, this can help add content to our conception of human nature. For example, we know that courage is a virtue that corrects for cowardice and foolhardiness; we know that temperance is a virtue that corrects for gluttony and abstemiousness. Grace, then, is love for the cowardly and the foolhardy, the intemperate, the immoderate, the stingy and the ostentatious, the boastful and the undignified – not because these qualities are *good*, but because they are *human*.¹¹⁶

By relying on this new repertoire of concrete qualities in informing our love of one another, the proposal that grace may be good to feel toward one's beloved is also a proposal to *notice* more about her – to attend to loved one in a realistic and truthful manner, and see more qualities about her that inform and reinforce one's affection. Spelling out full attention in this way helps further distinguish grace from forms of love justified by a person's bare humanity or personhood – such as the proposal offered by Velleman, or Kieran Setiya¹¹⁷, or just the familiar proposal offered in Christian thought that all human beings deserve love because we are God's creatures. As I discussed with Velleman above, these ideas seem pick out too thin a quality to serve as the right, or complete, object of full attention. Grace captures the phenomenological feel of attentive love on the part of the lover and the beloved – that it is affectionate and delighted attention to the actual qualities of the concrete individual.

Given the proposal I've offered, a question arises: are there limits on how bad a quality could be in order for it to fail to be a quality of humanity or human nature, and so not an appropriate quality for grace to be responsive to. What about sadism? What about deep racial hatred? Though I will not provide specific answers to specific questions, here too, an Aristotelian thought can further help us with the contours of this attitude and so the outlines of where the investigation would lead. In Book VII of the *Nicomachean Ethics*, Aristotle observes that along with *akrasia* and *vice*, another bad sort of state for a human being to be in is *thêriotês*, translated as the “bestial”, “beastly”, or “brutish”. Such states, he suggests intriguingly, are “less grave than vice, but more *frightening*”.¹¹⁸ Another

¹¹⁵ *Virtues and Vices*, 10.

¹¹⁶ This is similar in some respects to a point that Foot makes about our love of people who are intemperate, and imprudent but she notes that this is because one can be *wise* in one's intemperance and imprudence, and so, brings it under a good quality of a person after all. I think she is right to point this out about what it would take to be truly wise (and so, effectively rejecting Bernard Williams' claim that Aristotelianism must lead to a “depressing kind of moderation”). However, though I think we love the imprudence and intemperance of the wise, I think we can also love intemperance and imprudence. Bernard Williams, *Ethics and the Limits of Philosophy* (London, 1985), p. 36.

¹¹⁷ (Setiya 2014). What Setiya finds most dissatisfying with views of love that treat it as a rational response to positive qualities is that such a view yields the result that there may be some people who do not merit love (257).

¹¹⁸ NE 1150a37, my emphasis.

notable feature about this category of evaluation, as Pavlos Kontos notes, is that Aristotle's examples of beastly desires are "puzzling", in that – even keeping in mind cultural differences between the ancient Greeks and ourselves – they "hardly have any moral relevance"¹¹⁹. Indeed, they do not seem to exist along one dimension of *badness*, moral or otherwise: Aristotle includes examples of things that are terrible, such as ripping a child from a womb; but he also notes relatively benign things like plucking hairs and eating dirt.

Given this conceptual framework, there *are* limits to what qualities can count as human qualities, but the limits are not shaped to include just the good and exclude the bad. As Aristotle writes, "One sort of vice is human, and this is called simple vice; another sort is called vice with an added condition, and this is said to be bestial or diseased vice".¹²⁰ Human or "simple" vice, though bad in its own humanly way, is less *opaque* to us than bestial, or diseased "vice". And we react to a beastly person in a way that reflects this: we will not characterize or assess her behaviour in a morally robust way; she is not subject to the same attitudes, such as indignation, or resentment, or contempt, that are natural responses to human viciousness. Aristotle observes that we react to the beastly person with *fear*; Strawson would add other, non-reactive emotions such as revulsion, and pity – those we can feel from the objective stance. So, for each bad quality we are curious about, we will have to determine on which side of the line it falls – does it render someone "brutish", or humanly vicious? If it falls within the human, then – even if vicious – it may still be a candidate for the reactive attitudes that are available to us from the participant stance, the stance from which we may feel "the sort of love which two adults can sometimes be said to feel reciprocally, for one another".¹²¹

¹¹⁹ Kontos, "Non-Virtuous Intellectual States in Aristotle's Ethics", in *Oxford Studies in Ancient Philosophy*, XLVII, Brad Inwood (ed.), p. 224, fn. 39.

¹²⁰ (1149a17)

¹²¹ So: can extreme *immorality* or *evil* be the sort of thing that pushes a person out of the realm of the human and into the bestial? The question of whether or not to include evil into our conception of human nature (and so, whether evil could be a quality that grace is responsive to) is a question that must be settled by further moral and ethical reflection.

FROM RATIONALITY TO MORALITY: THE COLLECTIVE DEVELOPMENT OF PRACTICAL REASON IN KANT'S ANTHROPOLOGICAL WRITINGS

Olga Lenczewska

Introduction

Kant's "pragmatic anthropology"¹²² – the empirical study of moral philosophy – offers an account of our species' evolution into humans proper, i.e., into beings who not only possess the faculty of *reason*, but can make *full* or *mature* use of it. Generally speaking, such mature use of reason amounts to acting in accordance with the moral law (practical use of reason) and conducting a systematic study of nature (theoretical use of reason). In this paper I am interested in the former, i.e., in Kant's account of the way in which our species learned to make practical use of their reason in a developed, mature way. This account of practical reason's gradual development, which emerges from Kant's anthropological or empirical ethics, complements his more well-known transcendental ethics in many interesting ways (by, for instance, explaining the tight relation between our existence as a part of nature and our existence as independent of nature¹²³).

In an attempt to draw attention to its importance for Kant's moral philosophy, Allen Wood goes as far as to claim that Kant's ethics have an anthropological *basis*¹²⁴. I do not wish to adjudicate here which side of Kant's ethics – anthropological or transcendental – is more important or more fundamental for his philosophy. Rather, I am interested in the relationship between them, and in particular in examining how Kant's pragmatic anthropology prepares the way for his transcendental ethics by showing that our evolution into moral beings starts with our natural predisposition to develop capacities that will eventually *make us* moral – and, even though this evolution goes in some sense against nature, it at the same time is what makes us nature's "ultimate end". In this vein, my aim in this paper is to focus on the preliminary, anthropological elements of Kant's ethics, which can provide us with a valuable insight into the complex connection between our phenomenal and noumenal natures. I will argue that in the light of Kant's anthropological writings our attainment of the status of moral agents whose actions are governed by the moral law, and hence our attainment of the status of nature's "ultimate end", is the final stage of a sophisticated historical process of the development of our species' social and

¹²² Kant uses this term in the *Groundwork of the Metaphysics of Morals*, 4:388.

¹²³ As Kant writes in the "Conjectural Beginning of Human History", human history is nothing other than our species' "transition from the crudity of a merely animal creature into humanity, from the go-cart of instinct to the guidance of reason – in a word, from the guardianship of nature into the condition of freedom" [8:115].

¹²⁴ Wood 1991.

rational capacities – that, in other words, the development of the moral capacities of our species *depends on* the prior development of our social and rational capacities.

My essay has three parts. In the first part, I focus on the general role of reason in Kant’s systematic philosophy and contrast it with reason’s limited, initial use, portrayed in the “Conjectural Beginning of Human History”. In this essay, Kant conjectures that the process of the development of reason in our species started with the discovery of possessing a (not yet developed) faculty of reason. Reason’s initial, experimental use for creating desires independent from our natural instinct marked our attainment of negative freedom of will, i.e., our capacity for acting *rationally* (in accordance with hypothetical imperatives). However, at this point we were not yet capable of acting *morally* (in accordance with the categorical imperative).

The second part of my essay shows how the first use of reason in its capacity to act rationally created or awakened in us the propensity to evil and egoistic tendencies. The realization that desires can be manipulated prompted humans to enter into the condition of sociality for egoistic purposes: so that we can inflict injustice onto others and use them to satisfy our own desires. This was due to a characteristic feature of our psychological makeup qua social beings: “unsociable sociability”¹²⁵.

Reason was then further developed as we gradually improved the rules that govern our social order, perhaps each of us motivated to do so by his or her self-interest. The gradual improvement of our socio-political order can be described as the transition from reason’s egoism to pluralism: to just conditions of social co-existence enforced by a system of laws incentivizing people to comply with them despite their corrupt motivational system. This process is the focus on the third part of my essay, where I show how the gradual development of reason is, *inter alia*, the gradual development of just conditions of social co-existence. This mature use of reason, Kant claims, marks the attainment of the capacity to act morally (in accordance with the categorical imperative) in individuals and in the society as a whole.

1. The initial and the mature use of reason

In Kant’s philosophy in general, the faculty of reason is the highest of our mental powers and it is closely tied to the development of all other human predispositions. Without reason, neither our theoretical investigations into the laws of nature nor our practical capacity to set ends would be possible. Reason in its theoretical capacity produces regulative ideas that enable us to conduct a scientific investigation of the natural world¹²⁶. The necessity of such a regulative idea for our scientific investigation of nature becomes clear once we understand that, for Kant, organic nature (that is, all living beings, including ourselves)

¹²⁵ Cf. *Idea for a Universal History with a Cosmopolitan Aim*, Fourth Proposition.

¹²⁶ In §67 of the “Critique of the Power of Judgment”, Kant describes this faculty as producing “the [regulative] idea of the whole of nature as a system in accordance with the rule of ends, to which idea all of the mechanism of nature in accordance with principles of reason must now be subordinated” [5:379].

escapes the mechanistic-deterministic description. Furthermore, of all organic beings, human beings are the only species who through their faculty of reason can understand their place in nature and nature more generally. Thanks to this capacity, moreover, they can set and arrive at the “final end” of creation, which is a moral one: the realization of the “highest good”¹²⁷ on earth. This “highest good”, according to Stephen Engstrom, consists in “a maximum of virtue together with the happiness it conditions – that is, universal happiness proportioned to universal virtue”¹²⁸.

However, reason’s role in human life has not always been so wide in scope: in fact, one of Kant’s most interesting anthropological claims is that our capacity to make full use the faculty of reason requires reason’s development across time and across various individuals. In the “Idea for a Universal History with a Cosmopolitan Aim” (1784), Kant writes: “*In the human being (as the only rational creature on earth), those predispositions whose goal is the use of his reason were to develop completely only in the species, but not in the individual. (...) [R]eason itself does not operate instinctively, but rather needs attempts, practice and instruction in order gradually to progress from one stage of insight to another*”¹²⁹. Making full use the faculty of reason is therefore a project that is social and cross-generational. It follows that the mature rational capacities are learned capacities, even if the faculty of reason itself is innate.

But if the development of reason and its purposive capacities requires time, it must have started in some way, and its initial use must have looked differently to its full, mature use. In the “Universal History” essay, Kant describes the initial role of reason as that of “extending the rules and aims of the use of all its powers far beyond natural instinct, and it know[ing] no boundaries to its projects”¹³⁰. Those who possess reason, therefore, can think and do things they would not be able to think and do if their minds and actions were limited by their natural instinct. Being limited by natural instinct, as Kant clarifies in the essay “Conjectural Beginning of Human History” (1786), means being subjected to the determinate laws of nature by, for example, desiring only that which our biological instinct directs us to desire.

This, originally, was the condition in which all the living beings found themselves. Initially, Kant hypothesizes in the “Conjectural Beginning” essay, “Instinct, that *voice of God* which all animals obey, must alone have guided the novice. It allowed him a few things for nourishment, but forbade him [to use] others”¹³¹. Before the first use of reason, the primitive human’s relations to other objects and other living beings were fully determined by nature. Fully subjected to nature’s determinism, he had no free will and no capacity of

¹²⁷ Cf. *Critique of Practical Reason*, “The Dialectic of Pure Practical Reason”, Book II; *Critique of the Power of Judgment*, §84.

¹²⁸ Engstrom 1992, 750.

¹²⁹ *Universal History*, 8:18-19; original emphasis.

¹³⁰ *Ibidem*.

¹³¹ *Conjectural Beginning*, 8:111; original emphasis.

setting himself hypothetical imperatives. He could not be subject to any practical maxims, let alone the moral law.

At some point, however, Kant continues in the “Conjectural Beginning” essay, the primitive human recognized within himself the ability to notice other possibilities for acting than the ones given by instinct, and thus the ability to invent desires independently of his natural instinct. This – the creation of new desires that go against nature – was the first manifestation of our faculty of reason. What followed was the ability to compare options for acting, and finally the ability to *choose* how to act in order to satisfy a particular desire. Man therefore noticed his reason’s purposive capacity to determine his ends – to act rationally, according to self-legislated practical maxims. This “first attempt at a free choice”¹³², as Kant calls it, marked the beginning of our species’ emancipation from the animal kingdom and the tutelage of nature: “instead of the single objects of his desire to which instinct had up to now directed him, there opened up an infinity of them, and he did not know how to relate to the choice between them; and from this estate of freedom, once he had tasted it, it was nevertheless wholly impossible for him to turn back again to that of servitude (under the dominion of instinct)”¹³³. By having emancipated himself from the tutelage of nature, the human being attained negative freedom of will, which for Kant means independence from the mechanical coercion or from the determination by natural instinct¹³⁴.

2. Reason, egoism, and evil

The initial use of reason for setting for ourselves ends that go against nature had enormous consequences for our species’ further development. Most immediately, it was the cause of our feeling discontent for the first time and of the development of a vicious side of humans: the “evil” that hence became a part of our nature¹³⁵. This happened, Kant conjectures, because the pleasant feeling of being superior to other living beings that are unable to freely choose their desires and act independently of their instinct was soon replaced by the negative feelings of “anxiety and fright (...) concerning how he [the human being] (...) should deal with this newly discovered faculty. He stood, as it were, on the brink of an abyss”¹³⁶. From there, the realization that desires can be controlled (maximized, minimized, or changed altogether) coupled with the realization that the psychological makeup of all humans is the same led to the emergence of egoistic, evil desires, such as the desire to gain control over the others that in turn led to the development unequal and unjust social relations.

¹³² *Idem*, 8:112.

¹³³ *Ibidem*.

¹³⁴ Cf. *Critique of Pure Reason*, A564/B562; *Groundwork*, 445.

¹³⁵ Cf. *Conjectural Beginning*, 8:119.

¹³⁶ *Idem*, 8:112.

Kant's anthropological view of evil implies that its source is the condition of sociality¹³⁷. Within the condition of sociality, the human being saw himself as having to pay attention to others because they posed (or at least could potentially pose) a threat to the realization of his own desires and needs. Such conception of social relations prompted evil and egoistic behavior. According to Allen Wood and Sharon Anderson-Gold, evil "pertain[s] to us insofar as we are social beings" and is "closely bound up with our tendencies to compare ourselves with others and compete with them for self-worth"¹³⁸. The first use of reason and the entrance into the condition of sociality was therefore a fall, and our propensity to evil emerged almost simultaneously with, or shortly after, sociality.

This observation is most clearly formulated by Kant in the Fourth Proposition from the "Universal History" essay, where he writes: "*The means nature employs in order to bring about the development of all its predispositions is their antagonism in society, insofar as the latter is in the end the cause of their lawful order.* Here I understand by 'antagonism' the *unsociable sociability* of human beings (...). The human being has an inclination to become *socialized*, since in such a condition he feels himself as more a human being, i.e. feels the development of his natural predispositions. But he also has a great propensity to *individualize* (isolate) himself, because he simultaneously encounters in himself the unsociable property of willing to direct everything so as to get his own way and hence expects resistance everywhere"¹³⁹. Human beings activated their propensity to evil in their nature by entering into social relations with one another. And the type of evil that humans were capable of changed alongside the development of their social relations – for example, the emergence of authority relations led to political tyranny.

However, our propensity to evil is, paradoxically, not just the cause of numerous bad things that happened to mankind, but also the cause of everything good that happened to us. Most importantly, it is the cause of the development of a just socio-political order. This is because the gradual improvement of our socio-political order is a means to minimize our unsociable sociability and its large-scale manifestation – social antagonism. The development of social relations is closely connected to the attempts to minimize the evil individuals are capable of via socio-political means – attempts that, however, might have been motivated by individuals' self-interest and not by considerations over the well-being of others. As Kant writes in the "Universal History" essay, "Thanks be to nature (...) for the incompatibility, for the spiteful competitive vanity, for the insatiable desire to possess or even to dominate! For without them all the excellent natural predispositions in humanity would eternally slumber undeveloped"¹⁴⁰. The evolution of our socio-political order,

¹³⁷ For an interesting discussion challenging Wood's claim that the condition of sociality is the source and the necessary condition of humans' evil actions, see Jeanine Grenberg's "Social Dimensions of Kant's Conception of Radical Evil".

¹³⁸ Wood 1999, 287.

¹³⁹ *Universal History*, 8:20-21; original emphasis.

¹⁴⁰ *Idem*, 8:21.

therefore, is driven by the need to eradicate evil – and so by the very existence of evil¹⁴¹. Every human being, having noticed the egoism of his companions, was inclined to set up regulations or laws that enforce just and fair behavior, so that the self-interested others do not impede on the realization of his goals and desires. Therefore, the original reasons for developing a just socio-political order were egoistic or self-interested. (However, as we will see in the next part of this paper, these reasons change to altruistic or moral alongside the further development of our practical use of reason.) Without our natural propensity to evil that developed in the condition of sociality, we would not be capable of developing our socio-political order and our faculty of practical reason more generally.

What exactly, then, is the relation between reason and nature from the anthropological perspective? Kant's transcendental philosophy contains numerous assertions that "the end of the existence of nature itself must be sought beyond nature"¹⁴². The anthropological account of the relation between reason and nature mirrors this paradoxical idea present in his transcendental philosophy: that, even though nature made reason possible in the first place, the purposes of reason are at odds with nature. Since reason is the cause of unsociable sociability and social antagonism (features which in turn enable the development of all man's "excellent natural predispositions"¹⁴³), reason is a product of nature. More precisely, reason is nature's means to enable us to develop all our human predispositions and, as such, it is "the ground of perfectibility"¹⁴⁴. Since Kant understands "predisposition" to be "a feature of a living thing's nature that accounts for its developing in a certain way"¹⁴⁵, all our predispositions (such as that to evil) are natural. On the one hand, therefore, reason is a product of nature and nature's means to our evolution as a natural species; on the other hand, however, reason is the power of setting ends independently of (and even against) natural laws and natural instinct.

While the faculty of reason is innate to us, its development is a social enterprise. We can only develop reason through communicating with others, and the success of such an undertaking depends on the freedom to communicate. Our species' evolution is a social, or collective, project which cannot be undertaken by a single individual during his relatively short lifespan. In fact, when in the "Conjectural Beginning" essay Kant describes the first use of reason and the resulting activation of our propensity to evil as well as the tendency to compete with others for self-worth, he seems to be suggesting that the standpoint of a pre-social individual (i.e., an individual who has never been a part of a social order) who uses reason to create new desires for himself is *egoistic*, for such an individual is not capable of considering his own desires as equally important as the desires of others. Without fully-fledged sociality, an individual was not able to view other human beings as his equals, even though he was able to see himself as superior to animals that are driven

¹⁴¹ Kant's views on this matter were heavily influenced by Jean-Jacques Rousseau's "Discourse on the Origin of Inequality" of 1755.

¹⁴² *Critique of the Power of Judgment*, 5:378.

¹⁴³ *Universal History*, 8:21.

¹⁴⁴ Wood 1999, 301.

¹⁴⁵ *Idem*, 211.

solely by natural instinct. The first, pre-social use of reason led the human being to become a moral egoist, that is, someone who, as Kant writes in the “Anthropology from the Pragmatic Point of View” (1798), “locates the supreme determining ground of his will merely in his own happiness and what is useful to him, not in the thought of duty”¹⁴⁶. Moreover, Kant seems to suggest that humans entered the condition of sociality for egoistic reasons, for sociality started with self-interested individuals who wanted to exploit others to satisfy their own ends.

3. Reason, pluralism, and morality

Even though the first, pre-social use of reason led the human being to become a moral egoist, reason is not just the cause of this one feature of our moral psychology. On the contrary: once appropriately developed, reason can be also used to guide us to minimize our egoism and the evil in our nature. For any human being to see other human beings as equally valuable, he must first be part of an appropriately ordered society that promotes justice. How does this transition come about?

Even though the social condition was first constructed by self-interested individuals for egoistic reasons, it does not mean that a more developed, older society is still composed of people with these same motivational features. On the contrary: the social condition, Kant claims, has *changed* our way of perceiving one another. The development of a just socio-political order, initially triggered by self-interested individuals, happens alongside reason’s transition from egoism to “the disposition of not being occupied with oneself as the entire world, and conducting oneself as a cosmopolitan”¹⁴⁷ – a disposition Kant labels “pluralism”. Such development of our social capacities – using reason to guide us against the evil in our nature and egoism –, which happens after we have entered into the social condition, marks for Kant the attainment of the condition of enlightenment or maturity, in which we find ourselves once we have arrived at the disposition of pluralism. In the essay “An Answer to the Question: What is Enlightenment?” (1784), Kant defines the condition of enlightenment as “man’s emergence from his self-incurred immaturity [or tutelage]”¹⁴⁸, that is, emergence from the inability to use one’s understanding without someone else’s help due to lack of resolution or courage.

In a fully enlightened society guided by a pluralistic use of reason, every member would be free in an intellectual and a civil sense: no one would be treated as unequal by being subject to political, social, or religious oppression or coercion. As we read in the Fifth Proposition of the “Universal History” essay, “*The greatest problem for the human species, to which nature compels him, is the achievement of a civil society universally administering right*”¹⁴⁹. According to Allen Wood, Kant’s ethical thought in his anthropological writings

¹⁴⁶ *Anthropology from the Pragmatic Point of View*, 7:130.

¹⁴⁷ *Ibidem*.

¹⁴⁸ *An Answer to the Question: What is Enlightenment?*, 54.

¹⁴⁹ *Universal History*, 8:22.

focuses predominantly on “the human race’s collective, historical struggle to develop its rational faculties and then through them to combat the radical propensity to evil that alone made their development possible”¹⁵⁰.

During the transition from the condition of immaturity to the condition of maturity, moreover, we make a social, collective effort to moralize ourselves within the already existing society by combatting people’s propensity to evil and disciplining people’s will to follow the rational law of autonomy. In other words, we learn how to act in accordance with the categorical imperative: to treat moral duty as our strongest motivation to act. According to Allen Wood, “At every stage in the rational control of desire, what must be overcome is a dependence on the *natural* conditions of human *society*, which enslave rather than liberate the rational subject. Before there can be morality of autonomy, the subject must learn to act toward other subjects according to objective rules given through its own faculty of reason”¹⁵¹. The pluralistic use of reason that leads to the formation of a just civil society composed of free and equal human beings is also responsible for the emergence of moral law and actions guided by it: “through [a continuous] process of enlightenment”, Kant writes in the Fourth Proposition of the “Universal History” essay, “a beginning is made toward the foundation of a mode of thought which can with time transform the rude natural predisposition to make moral distinctions into determinate practical principles and hence transform a *pathologically* compelled agreement to form a society finally into a *moral* whole”¹⁵². Therefore, morality is developed during this process of constructing a growingly more just society. It is a socially and historically developed product of our mature use of reason, and, as such, the end product of the rationalization of our lives. Our reasons for coexisting in the condition of sociality change from egoistic (self-interested) to altruistic (moral) alongside practical reason’s development.

Reason’s mature practical use, therefore, consists not just in the ability to set ourselves ends independent of our natural instinct (which was the original use of reason described in the “Conjectural Beginning” essay), but also in constructing the moral law and the ability to act purposefully in accordance with the categorical imperative. Anthropologically speaking, obeying the moral law and fulfilling one’s duty to do so would mean having successfully eradicated the evil in one’s nature. Duty is a positive incentive, whereas the propensity to evil gives us motivation to act on other, non-moral incentives. The more successful one is in obeying the moral law, the less of a moral egoist one is.

Of course, the sheer *ability* to act in accordance with the categorical imperative – to feel a stronger incentive to act on moral duty than to act for other reasons (on non-moral “inclinations”) – is not the same as actually *acting* in accordance with the categorical imperative. Even beings with a substantively developed moral capacity may occasionally fail to act on duty. It is precisely this distinction – between the ability to act morally and

¹⁵⁰ Wood 1999, 296.

¹⁵¹ *Idem*, 240.

¹⁵² *Universal History*, 8:21.

actually acting morally – what makes our actions either morally evil or morally good¹⁵³. And it is precisely this distinction, but scaled up so as to concern every citizen of a state, that led Kant to distinguish between “a juridico-civil state” and “an ethical state” in his later anthropological thought from the “Religion within the Boundaries of Mere Reason” (1793).

Because our sheer ability to act in accordance with the categorical imperative does not guarantee that we will always do so, a society composed of morally imperfect citizens that aspires to be a just society – “a juridico-civil state”¹⁵⁴ – must have legal means of enforcing justice. In other words, the fact that some people are morally corrupt necessitates a coercive system of rules, laws, and punishments, implemented on a state level¹⁵⁵. Kant talks about the motivational problem in Part I of the “Religion”. He gives examples of motivational failures to act purely on duty by distinguishing three degrees of such failures: “frailty” (when we sometimes prefer non-moral incentives to moral ones), “impurity” (when the disposition to do evil is a part of the motivational structure of actions that accord with duty), and “depravity” (when we do our duty only when we also have sufficient non-moral incentives to do it)¹⁵⁶. This discussion in the “Religion” occurs in connection with Kant’s exposition of our species’ third and final predisposition¹⁵⁷ – the predisposition to *personality* – which is gradually realized after we have gained the capacity to act in accordance with the moral law. This capacity, amounting to the feeling of respect toward the moral law, figures in our motivational system alongside other, non-moral, reasons for actions: “The predisposition to personality is the susceptibility to respect for the moral law *as of itself a sufficient incentive to the power of choice*. This susceptibility to simple respect for the moral law within us would thus be the moral feeling, which by itself does not yet constitute an end of the natural predisposition but only insofar as it is an incentive of the power of choice”¹⁵⁸.

It is our predisposition to personality that enables us to avoid the motivational failures of frailty, impurity, and depravity. This is because the predisposition to personality amounts to our developing of a “good character” thanks to which we are in a position freely to

¹⁵³ Kant’s detailed and interesting discussion of what makes someone good or evil may be found in Part I of the “Religion” (especially 6:36).

¹⁵⁴ *Religion*, 6:95.

¹⁵⁵ Note that the attainment of the ability to act in accordance with the categorical imperative changes conceptually what it means to act in an evil way. The anthropological history of evil that Kant seems to portray requires, however, a separate and much more detailed discussion than I am able to provide in this paper.

¹⁵⁶ Cf. *Religion*, 6:29-30, 6:38.

¹⁵⁷ As Kant writes in Part 1 of the “Religion”, there are three predispositions that belong to human nature: “1. The predisposition to the *animality* of the human being, as a living being; 2. To the *humanity* in him, as a living and at the same time rational being; 3. To his *personality*, as a rational and at the same time responsible being” [6:26]. We may call the predisposition to personality a final predisposition because it gets developed through the processes of civilization and moralization, and because the two other predispositions are the necessary conditions for its development.

¹⁵⁸ *Religion*, 6:27.

choose acting on the moral feeling rather than on the non-moral inclinations. Acting in accordance with the moral law, Kant writes in the “Religion”, “is possible only because the free power of choice incorporates moral feeling into its maxim: so a power of choice so constituted is a good character, and this character, as in general every character of the free power of choice, is something that can only be acquired”¹⁵⁹.

But how is a good character acquired? An answer to this question may be found in §83 of the “Critique of the Power of Judgment”. According to it, our moralization – the moral education or the cultivation of one’s moral capacities – takes place through or alongside the gradual and collective cultivation of our species’ other (cognitive and aesthetic) capacities¹⁶⁰. For Kant, the fact that moral education is possible means that successful construction of just laws and rules in a pluralistic society – a construction of a “juridico-civil state” – is not the final stage of our reason’s practical capacities. What is further needed is the gradual and collective cultivation or development of our species’ cognitive and aesthetic capacities – capacities through which we exercise and train our faculties. Through the gradual education of their moral sentiments, which ought to happen via the development of culture, arts, and sciences, the members of a just society could eventually become perfectly moral, which would amount to always choosing to act on duty and not on non-moral incentives.

§83 of the “Critique of the Power of Judgment” therefore connects our species’ natural predisposition to develop various forms of culture with a central end of pure practical reason: the realization of a perfectly just society. Only by pursuing our natural predispositions to create art and to conduct a scientific investigation of nature will we be able to give objective reality to the ideal of a perfectly just society. Kant concludes §83 by saying: “Beautiful arts and sciences, which by means of a universally communicable pleasure and an elegance and refinement make human beings, if not morally better, at least better mannered for society, very much reduce the tyranny of sensible tendencies, and prepare humans for a sovereignty in which reason alone shall have power”¹⁶¹. For us to be “sovereign” from “sensible tendencies” and to be empowered by “reason alone” would presumably mean to be able to exercise our reason’s capacities to their fullest – indeed, by having our actions always guided by the categorical imperative and not be led astray by self-interested non-moral incentives.

This ideal of perfectly moral citizens who always choose to act on duty is an ideal of the formation of an “ethical community”¹⁶², as Kant writes in the “Religion”. Such a community – also called an “ethical state” in direct contrast with the “juridico-political

¹⁵⁹ *Idem*, 6:27.

¹⁶⁰ Following Rousseau’s sentimentalist claims about the role of education in society, Kant sees a tight connection between general education of citizens and the attainment of one’s moral and political freedom. For example, Kant notes on his copy of his own pre-critical work “Observations on the Feeling of the Beautiful and Sublime”: “The chief intention of Rousseau is that education be free and also make a free human being” [*Notes and Fragments*, note 46].

¹⁶¹ *Critique of the Power of Judgment*, 5:433.

¹⁶² Cf. *Religion*, 6:94-5.

state” – would only need an ethical legislation (one that is freely accepted by all its members), not a legislation that is coercive in any way. Kant contrasts the two types of states in the following way: “A *juridico-civil* (political) *state* is the relation of human beings to each other inasmuch as they stand jointly under *public juridical laws* (which are all coercive laws). An *ethico-civil* state is one in which they are united under laws without being coerced, i.e. under *laws of virtue* alone”¹⁶³. In the latter, every member’s decision to act in a just way must be made freely and autonomously if it is also to count as a morally good decision. If, by contrast, one acts in a just way because of heteronomous incentives (such as being afraid of the legal consequences of doing otherwise), then one’s action, even if seemingly or empirically good, is intelligibly evil and thus does not count morally good. To overcome evil, one must act freely and autonomously, through self-imposed incentives¹⁶⁴. In an ethical community, human beings would therefore willingly treat one another as equals and would freely to subject themselves to the moral law. This community would be united through the concept of a moral final end (the “highest good”) pursued by all the members as a collective end: “[the] highest good will not be brought about solely through the striving of one individual person for his own moral perfection but requires rather a union of such persons into a whole toward that very end”¹⁶⁵. This is where the anthropological branch and the transcendental branch of Kant’s moral philosophy come together.

I will now, by way of conclusion, attempt to link Kant’s anthropological account of our species’ development, presented above, with the transcendental side of his moral philosophy, and particularly with the teleological conceptions of the ultimate end of nature and the final end of all creation.

Kant believes that a successful formation of the ethical community as a perfectly just civil society (and, subsequently, its international version) will mean that our human natural predispositions have been developed to their fullest. This, as we have seen, is the key point of §83 of the “Critique of the Power of Judgment”. It is also the key point of the Eight Proposition of the “Universal History” essay, which states: “*One can regard the history of the human species in the large as the completion of a hidden plan of nature to bring about an inwardly and, to this end, also an externally perfect state constitution, as the only condition in which it can fully develop all its predispositions in humanity*”¹⁶⁶. What Kant means by saying that the development of such a socio-political order is “a hidden plan of nature” is that the development of our human natural capacities (cultural development) is

¹⁶³ *Idem*, 6:95; original emphasis.

¹⁶⁴ Cf. *Religion*, 6:36-37, and in particular: “[In a] reversal of incentives through a human being’s maxim contrary to the moral order, actions can still turn out to be as much in conformity to the law as if they had originated from true principles (...). The empirical character is then good but the intelligible character is still evil. (...) [T]here is the human being a natural propensity to evil (...). Yet it must equally be possible to *overcome* this evil, for it is found in the human being as acting freely”.

¹⁶⁵ *Religion*, 6:97-98.

¹⁶⁶ *Universal History*, 8:27.

the realization of the ultimate end¹⁶⁷ of nature. And nature must have an ultimate end (i.e., it must be regarded by us as having one), for otherwise we could not conceive of it as a finally unified teleological system – but such a conception is required for us to be able to conduct a systematic investigation of nature and its laws. Therefore by fully realizing our natural predispositions, which amounts to fully developing our faculty of reason and extending our capacities to act as free beings, we can become the ultimate end of nature. In §83 of the “Critique of the Power of Judgment”, Kant writes: “As the sole being on earth who has reason, and thus a capacity to set voluntary ends for himself, he [the human being] is certainly the titular lord of nature, and, if nature is regarded as a teleological system, then it is his vocation to be the ultimate end of nature”¹⁶⁸. Our natural vocation, therefore, is to become the ultimate end of nature – to develop our natural human capacities to their fullest.

Given that nature must have an ultimate end and that we are a part of this nature, we need to think of the history of our species as being subordinate to a teleological ideal, whose realization must moreover be regarded as *possible*. However, as Kant makes clear at the beginning of the “Universal History” essay, such systematicity can never be discovered by us within it, but can merely imposed on it as an approximation: history, as a discipline that is concerned with giving an account of human actions, “allows us to hope from it that if it considers the play of the freedom of the human will *in the large*, it can discover within it a regular course; and that in this way what meets the eye in individual subjects as confused and irregular yet in the whole species can be recognized as a steadily progressing though slow development of its original predispositions”¹⁶⁹. Therefore, the teleological ideal to which our history is subordinate is a *regulative* ideal.

As Kant further specifies in §83 of the “Critique of the Power of Judgment”, the human being can only become the ultimate end of nature “subject to the condition that he has the understanding and the will to give to nature and to himself a relation to an end that can be sufficient for itself independently of nature, which can thus be a final end”¹⁷⁰. That is, becoming the ultimate end of nature can only take place if we succeed in self-imposing the moral law on ourselves, and thereby subordinating our history to something higher than natural laws – to a perfect socio-political order (an ethical community) governed by appropriate self-imposed moral rules. In other words, we can only become the ultimate end of nature by realizing the final end of all creation¹⁷¹, which is a moral end: the “highest

¹⁶⁷ The ultimate end is a natural end that completes an aggregate of lower natural ends, or a regulative idea that we employ in order to be able to think of a teleological system as a unified whole.

¹⁶⁸ *Critique of the Power of Judgment*, 5:431.

¹⁶⁹ *Universal History*, 8:17; original emphasis.

¹⁷⁰ *Critique of the Power of Judgment*, 5:431.

¹⁷¹ A final end, in general, is an end that does not need another end as the condition of its possibility. And the final end of all creation, as Kant explains in §82 and §84 of the “Critique of the Power of Judgment”, must be something outside the order of nature (“if we go through the whole of nature, we do not find in it, as nature, any being that can claim the privilege of being the final end of creation” [5:435]), for the final end is unconditioned but everything within the order of nature is conditioned (“the final end cannot be an end that nature would be sufficient to produce in accordance with its idea, because it is unconditioned. For there is

good”. Human beings, as the only species possessing reason and thus capable of setting ends independent of nature, are the only species that can become the final end of all creation: “only in the human being, although in him only as a subject of morality, is unconditional legislation with regard to ends to be found, which therefore makes him alone capable of being a final end, to which the whole of nature is teleologically subordinated”¹⁷².

The realization of the final end of all creation, i.e., the full development our moral capacities and the formation of a perfectly just society, would conclude the realization of ultimate end of nature and thus the realization of our species’ vocation or, to put it in anthropological terms, of our species’ natural predispositions. Without having subjected our autonomous will to the moral law and so without having constructed an ethical community, we cannot become the final end of all creation. But to be able to construct an ethical community in the first place, we must first realize the ultimate end of nature, i.e., develop our natural predispositions to civilization, culture, science, and artistic expression. The cultivation of our moral capacities happens alongside the cultivation of our other capacities: cognitive and aesthetic. Therefore, the development of the *moral* capacities of our species (and thus the creation of a perfectly just society: an ethical community) is the final stage of our species’ natural development – a stage which itself transcends the phenomenal realm of nature. This development, as was my goal to show in this essay, would not be possible without the prior development of our rational and social capacities – a development which is necessarily a collective enterprise.

Bibliography

Engstrom, Stephen (1992), “The Concept of the Highest Good in Kant's Moral Theory”, in *Philosophy and Phenomenological Research* 52/4: 747-780.

Grenberg, Jeanine M. (2009), “Social Dimensions of Kant’s Conception of Radical Evil”, in *Kant’s Anatomy of Evil*, eds. S. Anderson-Gold & P. Muchnik. Cambridge: Cambridge University Press: 173-194.

Herman, Barbara (1981), “On the Value of Acting from the Motive of Duty”, in *The Philosophical Review* 90/3: 359-82.

nothing in nature (as a sensible being) the determining ground of which, itself found in nature, is not always in turn conditioned” [5:434]).

¹⁷² *Critique of the Power of Judgment*, 5:435-6.

Kant, Immanuel (1991), *An Answer to the Question: What is Enlightenment?*, in *Political Writings*, trans. H. B. Nisbet. Cambridge: Cambridge University Press.

Kant, Immanuel (2006), *Anthropology from a Pragmatic Point of View*, trans. R. B. Loudon. Cambridge: Cambridge University Press.

Kant, Immanuel (2007), *Conjectural Beginning of Human History*, in *Anthropology, History, and Education*, trans. A. Wood. Cambridge: Cambridge University Press.

Kant, Immanuel (2000), *Critique of the Power of Judgment*, trans. P. Guyer & E. Matthews. Cambridge: Cambridge University Press.

Kant, Immanuel (2005), *Notes and Fragments*, trans. C. Bowman, P. Guyer & F. Rauscher. Cambridge: Cambridge University Press.

Kant, Immanuel (2007), *Idea for a Universal History with a Cosmopolitan Aim*, in *Anthropology, History, and Education*, trans. A. Wood. Cambridge: Cambridge University Press.

Kant, Immanuel (1998), *Religion within the Boundaries of Mere Reason*, trans. A. Wood & G. di Giovanni. Cambridge: Cambridge University Press.

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

Wood, Allen (1991), "Unsociable Sociability. The Anthropological Basis for Kantian Ethics", in *Philosophical Topics* 19/1: 325-351.

On Being a Rational Animal and Public Reasons

Troy Seagraves

Naturalism according to Hursthouse

Hursthouse's main claim is that a good person is equivalent to a good human organism. We evaluate an organism according to standards of what is characteristic of its species. This is (roughly) the nature of the species. For instance, wolves hunt in packs. If we found a wolf that refused to hunt with other wolves, we would call it a defective wolf. This is because wolves are social animals and are evaluated as such. Hursthouse notes how we evaluate social forms of life:

...A good social animal... is one that is well fitted or endowed with respect to (i) its parts, (ii) its operations, (iii) its actions, and (iv) its desires and emotions; whether it is thus well fitted or endowed is determined by whether these four aspects well serve (1) its individual survival, (2) the continuance of its species, (3) its characteristic freedom from pain and characteristic enjoyment, and (4) the good functioning of its social group—in the ways characteristic of the species.¹⁷³

It is important to note that the four ends (1-4) are the ends towards which social animals strive. A species' characteristic way of living is just a way to achieve those ends. So, we can evaluate defects based upon not only how an organism deviates from its characteristic form of life (its nature), but also how an organism fails to achieve the four ends towards which that the nature works. A characteristic form or way is a general (but not statistical) tendency in the species. That is, an organism under ideal conditions acts according to the characteristic way. For instance, geese fly south for the winter when we do not confine them to cages. We can say that a social animal's nature is its characteristic way of endorsing the four ends of social animals.

Human evaluations follow along the same basic lines as above: we evaluate a human organism according to how well it abides by the characteristic way its species promotes the four ends. Humans differ in that they are rational animals and their characteristic way of achieving the four ends is by expressing rationality, or a "rational way." Hursthouse writes: "A 'rational way' is any way that we can rightly see as good, as something we have reason to do."¹⁷⁴ The "rational way" is acting on reasons that are, in fact, the right reasons. This is done via rationality, a part of our nature, which is expressed in a multitude of ways and allows us to evaluate to what extent our common ways of doing things is good. That is, if we see a trend in our species that does not achieve the four ends, then we can decide to change it.¹⁷⁵ Rationality has this power of reflection and evaluation of reasons for action. The virtues for humans are those traits that are part of the "rational way" and achieve the four ends towards which our species works.

¹⁷³ Hursthouse, (1999), 155.

¹⁷⁴ Ibid. 222.

¹⁷⁵ Ibid. 220-1.

II. The Normativity Problem

One of the most common objections to Hursthouse's view questions the normativity of natures. Consider the fact that a given organism has a certain nature. What prevents it from rejecting its nature, rejecting the very way Hursthouse says we evaluate organisms? Several authors claim that humans have this ability to reject their nature. Summarizing the literature succinctly, Micah Lott gives what he calls the authority-of-nature challenge:

Our rationality enables us to step back and ask whether we should live in the way that is 'naturally good' for us. And this question can only be properly answered with a reason—i.e., with a point about why it is good to live one way or another. An answer that appeals simply to our 'nature' is of the wrong sort, since it just re-asserts that this is the way things are with us, rather than giving an answer to the question of why we should embrace this way or depart from it when attempting to live and act well.¹⁷⁶

If human nature is the characteristic way humans achieve the four ends of social animals, then rationality seems able to scrutinize human nature. Specifically, rationality allows us to evaluate how we generally do things. Take John McDowell's wolf. He notes:

Just how convincing a grounding for the appeal of ethical considerations to reason is available anyway, from the claim that human beings need the virtues if their life is to go well? Would this claim be like the claim that wolves need a certain sort of cooperativeness if their life is to go well? Suppose some wolves acquire reason...Having acquired reason, he can contemplate alternatives; he can step back from his natural impulse and direct critical scrutiny at it.¹⁷⁷

McDowell's wolf uses rationality to reject one of its characteristic ways of doing things (being cooperative). The role of reason in the example is crucial. The sting of the objection comes from the rational, reflective ability to reject methods and suggests that this ability seems to allow a rejection of one's nature altogether. If we were to convince the wolf to be cooperative, we cannot appeal to the nature of wolves. The rational wolf would be unconvinced if we merely told her that wolves characteristically are cooperative. This ineffective approach, however, seems to be the route Hursthouse has left us. If one's nature is the standard for goodness, then it should be sufficient to show an organism its nature to show it how to be good.

Recall that Hursthouse says that humans are evaluated according to their use of rationality to achieve the four ends of social animals. Call this our *social character* (so, Hursthouse's claim would be that an organism has good social character when she promotes her species' ends). McDowell's objection may be interpreted to question why we should accept the four ends as standards of evaluation. He questions why the virtues may be needed to define a good life, and the virtues are those rationally endorsed methods that promote the four ends. McDowell's objection then joins David Copp and David Sobel who

¹⁷⁶ Lott, (2013), 767.

¹⁷⁷ McDowell, (1998), 153.

express a similar objection to Hursthouse: “How can Hursthouse reject the thought that nature determines how humans should be yet think that the same considerations that grounded the four ends in plants and animals also ground the normative status of the four ends for humans?”¹⁷⁸ McDowell, Copp, and Sobel’s objection (hereafter, the MCS objection) seems serious: We can use our rational way to reject our nature as social animals, reject the four ends which Hursthouse uses to ground human evaluation. In the next sections, I aim to argue that we can avoid the MCS objection by appealing to the nature of reasons.

III. Rationality and Social Agents

To avoid the MCS objection, we will need to discuss in terms of social animals. Hursthouse’s use of “social animals” is technical, but it is unclear how it jibes with our normal use of the term. To avoid confusion, we can use “social agents” to mean the technical version as well as the development to come. The strategy here is to investigate how social agents use reasons.

In recent literature, reasons have been defended as *public*.¹⁷⁹ When I say that reasons are public, I mean that reasons are shaped and disciplined by public discourse. Call a purported reason-giving notion a *claim*.¹⁸⁰ Claims are proto-reasons. That is, they are the things that possibly become reasons. The first step in this take on the publicity of reasons is to show that reasons and claims are molded by others. We are susceptible to the input of others when in public discourse. We cannot simply ignore what others say, but the moment we hear the input of others our own thoughts change.¹⁸¹ That is, the content of the claims may not change in discourse, but those claims will have a different flavor or hue in the mind. In many cases, the context is what changes. We come to have the same claims, but they are cast in a different light due to the context of recalcitrant claims. Suppose I, on a whim, take a plane to Austria. If you object to my decision (and I hear it) I cannot ignore this objection. It is an unanswered objection. I can still take the plane, but my whim will not be the same. The whim appears in my mind *next to the unanswered objection*, in the shadow of this threat. This new content changes the hue of our thoughts in that, next to unanswered objections, we cannot help but be less confident in our motivations for acting.

To take the step towards necessarily public vindication, we need to show that there is no private vindication. This will help show that without public discourse, we do not have reasons. Vindication is not when some hidden nature of our claims is suddenly revealed to us. Vindication is how reasons are formed from claims. Moreover, reasons are those that have been molded by others. Notice that once we have a claim for some action or way of life, we need to know that it is a reason and not just an impulse. This is not to say that impulses are never reasons, but to *become* reasons (rather than merely claims) they need to pass the test of public discourse. This test is a cooperative project. In our interaction with

¹⁷⁸ Copp and Sobel (2004), 540.

¹⁷⁹ Korsgaard (1996) and Lebar (2008).

¹⁸⁰ Darwall, (2001). 146.

¹⁸¹ Korsgaard, (1996), 139-140.

others, we give and receive claims. These claims, once vindicated, become reasons as well. Rationality demands some type of vindication, a criterion of correctness for our reasons.

When we have rationality, we have the capacity to recognize there is a correct way to do something. Our having reasons for doing something at all seems to be our recognition that there is a right way of doing it. It is this recognition (and the process that accompanies it) that makes private vindication not possible. Let us imagine what private vindication would look like. Perhaps it is noticing the claim in one's own mind, then reflecting on whatever else is in said mind, attempting to vindicate the claim into a reason. This introduces a process where one part of the mind is reformed to cohere with the other parts. This process continues until the entire mind coheres with itself. After full coherence is established, it is still unclear what is shown. Our quest for the right reasons does not stop there. We can question further whether the fully coherent mind is right, and this question requires a response. To check the mind's validity, it seems we are left with reflecting on those parts of the mind we have already reflected upon. Coherence alone is too weak of a criterion since, depending on the reformation process, just about anything can pass the test of coherence so long as the rest of the mind cooperates. So, anything can be correct.¹⁸² Such a conclusion rooks the concept of "correct" of any meaning. Private vindication negates the need for correctness which is what we said rationality pushes us to pursue. If we are to vindicate something like reasons, we must search for such vindication outside of ourselves.

Someone may object, insisting that we can vindicate a claim privately. The best case for this objection would be when we are alone. Clearly, we are not always in public discourse and it seems we can have reasons even when not in discussion with others. We certainly can, and my view does not bar the lonely from having reasons. To illustrate how, consider the case of playing chess. It is my move and I immediately see that I should move the knight or the bishop. When I deliberate about which move to make, I do not speak with my opponent or consult those around me. This, it seems, is private vindication, but it is not. Think of how I *actually* deliberate on which move is better: I recall previous games, theory, lessons, and principles I have gathered over time. My deliberation is reliant on persons and processes that are external to myself. The decision I make is from a reason so long as I have *already* engaged in public discourse; I appeal to a public entity.

Now, consider a chess prodigy playing her first game. She does not have these resources so the move she makes is based on a claim (a whim or a guess). Her move may have been brilliant, but she does not have the backing that *ensures* that her proposed move is a good one. Her seeing a good move does not, on its own, give her a reason to move a piece. She can still question the validity of the seeing and, without a public entity of some sort, this question will remain unanswered. Suppose further that the prodigy plays several games but remains taciturn in her interaction with others. She may have claims that succeed in that they lead to wins, but these are still not reasons. Recall that vindication may change nothing but the context of the claim or hue of the mind. Until unanswered objections are

¹⁸² This is similar to Wittgenstein's Private Language Argument. Wittgenstein, (1968), §265.

met, the hue is still negative. The seeing is still questionable and we need external parties, a public entity, to ensure that the prodigy has actually made good moves.

Now, we should become clearer on the public side. So far, I have been using “public discourse” as a catch-all term for whatever the public part of vindication would be. To be clear, there is some appeal to how the term “public discourse” is traditionally used. That is, discussing with other people is central to the notion.¹⁸³ This discussion is aimed at finding the right reason to do something.¹⁸⁴ Since this is the case, we can think of public discourse as a mechanism to combat *the skeptic*. This phantom character questions reasons we have (or at least those at hand). The skeptic comes in forms of actual objections or internal doubt that we could be wrong. On this view, having rationality means one recognizes the importance of answering the skeptic. Combatting the skeptic is quite difficult to do and can only be done by constantly refining, shaping, and disciplining reasons. We may, then, question how many and what type of people we ought to discuss with if we are to vindicate. In short, we need whatever it takes to allow us to confidently quiet the skeptic. We do not need to *convince* the skeptic, but we need to create a defensible reason where the skeptic’s questions are no longer troubling or vindicate some reason that justifies ignoring the skeptic. Doing so should quell any relevant suspicion that you could be wrong. Consider how John Stuart Mill is traditionally read on public discourse and truth. Answering skeptics and objections is key to understanding the truth. We must, even in the absence of present skeptics, consider skeptical objections for the sake of vindication.¹⁸⁵ So, there is no formula for how to conduct public discourse¹⁸⁶ but the point is that it is our way to answer objections, find our best claims, and vindicate claims into reasons.

Having reasons admits to the possibility of a correct way of doing things and plumbing this is a cooperative project. We seek out and become involved with others in order to construct reasons. This is the social side of our rationality. Recall that Hursthouse discusses the “rational way” that is characteristic of human beings as any way that we can rightly see as good. Our recent discussion of the publicity of reasons sheds some light on what is going on here. If reasons are public in the way I have sketched, then our cooperative vindication is expressive of our need for the right reasons. We are constantly striving towards the “rational way” that is characteristic of humans. The search for vindication via public discourse is just our attempts to get at the right way. We do this by being social, by engaging with the public.

IV. The Public Machine

¹⁸³ This structure is very congenial with recent work on virtue. Julie Annas has provide an account of virtues as habits that obtain from the giving and taking of reasons. See Annas, (2011), chapter 2.

¹⁸⁴ Some may worry about how exactly this discourse gets off the ground. We may need to agree on something to not merely talk past each other, stultifying this quest for the right reasons. This is plausibly correct and does not threaten this project. See MacIntyre, (2016) 56-7.

¹⁸⁵ Mill, *On Liberty*, 243ff.

¹⁸⁶ We can, however, look to the best ways to conduct public discourse and vindicate reasons. This metaethical framework fits in intellectual virtues here. See Zagzebski, (1996) and Baehr, (2011).

We can now tackle the four ends Hursthouse claims social agents must promote. With the publicity of reasons, we saw the public nature of rationality. What we need now is an account of how the publicity of reasons establishes our social character. Recall that the four ends by which a social agent is evaluated according to Hursthouse are: (1) its individual survival, (2) the continuance of its species, (3) its characteristic freedom from pain and characteristic enjoyment, and (4) the good functioning of its social group. Going from the publicity of reasons to these four ends is tricky, so an analogy may be helpful: Think of the publicity of reasons as a machine. This machine is whatever it is that public vindication involves. A social agent, having reasons, would want to maintain this machine. To be clear, this does not come from having the right reasons. Public vindication does not guarantee acquiring the right reasons. From having reasons at all, the social agent must endorse the machine. The strategy, then, is to show how each of the four ends is part of this maintenance.

First, we have the social agent's individual survival. When tackling this, we must remind ourselves that, as Hursthouse avers, the survival of individual organisms is something that more than social agents must promote. Even a nonsocial agent that haphazardly cares for itself and leads to its own demise, it is a bad agent. Now, maintenance of the machine includes anything that is entailed by the nature of reasons. If something is taking part in the machine, then it has reasons (or at least claims) to act or believe. By merely having such reasons, an organism has (defeasible) reason to promote its own survival. An organism may have reason to act courageously, but this includes some reason to survive since acting ceases upon the demise of the organism. This does not, however, preclude reasons to sacrifice oneself. Since any reason can be revised in the machine with new information, there are cases of legitimate self-sacrifice.

Next, the survival of an organism's species utilizes the social aspect of the machine. In discussing this aspect, we will not be aiming to show why the social agent should procreate. Such a reason will most likely come out of public discourse. The subject here is the social agent acting to assist the flourishing of future generations. To this end, we may part from Hursthouse. We come facing two facts about the public machine: (1) the "raw materials" of the machine are individuals and (2) those individuals are (contingently) those of your species. From these two facts, we have some need for a social agent to endorse the survival of its contemporary species members. To deny the machine these materials is to reject the operation of the machine thus contravening the formation of reasons. So, a social agent has some reason to promote the survival of its species.

Now that we have this reason, we need to consider how it includes the right amount of (future) people. As we saw from the nature of public discourse, the group that one needs for the public machine is not merely those around you. We need at least enough to satisfy the quest for the right reasons. This extends back in time. As we know, we can learn from those who came before us. This also extends contemporaneously as I discover cultures of today that hold different views than I do. With this information alone, I must investigate the validity of the views at hand. Further, if I want the right reasons, then I want others to follow the right reasons as well. Rationality forces us to care about the right reasons, but this does not mean that it forces us to care only for *my having* the right reasons. Rationality does not discriminate between agents. If something is the *right* reason, then I do not deny

it to you. Think of actual public discourse. When we see the need for fraternal correction, we do not remain laconic, but we seek to correct those who are wrong. If public discourse was only about *my* gaining the right reasons, then I would have no business discussing further once I have vindication. Fraternal correction is my business because rationality forces me to care about the right reasons regardless of who has them. So, if I have reason to pursue the right reasons, then I also have reason to facilitate in others pursuing the right reasons. This includes future generations as they are beings that are capable of having the right reasons *and this is all I need* to care about them. That is, since future generations will be capable of having reasons, I will want them to have the right reasons. This allows the public machine to extend forwards in time as we have reason to promote the flourishing of future generations.

The next topic, the characteristic freedom from pain and enjoyment, uses the analogy of a public machine in a different way from above. We try to avoid pain but also, we try to feel enjoyment. This merits some comment before it can be included in the analogy. Clearly, we do not try to avoid every kind of pain or displeasure. Consider the case of running. Running is clearly painful in some respect (muscles burn, aches, etc.) but runners are admirable for their diligence and fitness. Running is quite different from forcibly straining your breathing. While running has acceptable pain, forcibly straining your breathing is *harmful*. In running, your breathing may become strained but this instance of it serves the function of running. Our characteristic freedom from pain needs to capture this distinction. Likewise, our characteristic enjoyment needs to differentiate between those who enjoy meaningful activities from drug addicts stuck in a euphoric stupor.

From the above comments, we can see a sort of moderation captured in our characteristic freedom from pain and enjoyment. Now, think of the publicity of reasons. The operators of the public machine are people who experience pain and pleasure. Our actions keep the machine going but this also means that we acknowledge the factors that keep the machine from operating well. Once acknowledged, we can work to avoid such factors. Like an engine that needs fuel but should be kept from flooding, we can look at pain and pleasure as factors that affect our operating abilities. With too much pain (or pain of certain types) we decrease the efficiency of the machine. Likewise, too much pleasure keeps our reasoning capacities below their optimal level. Endorsing the public machine also calls for endorsing our characteristic freedom from pain and enjoyment because doing so keeps the public machine operating in the way it should. As I have expounded on this part pleasure does not come out to be intrinsically good and pain does not come out to be intrinsically bad. They both play a part in the maintenance of the machine and are valuable in certain types and quantities. If we accept what I have argued so far, then this should not be a problem since the aim is to find the right reasons.

Lastly, the good functioning of the social group can be seen as general maintenance of the public machine. The social agent must promote this because, as we should recall, the public machine is the very thing that produces vindicated reasons for the social agent. Taken narrowly, the good functioning of a social group includes avoiding unnecessary conflicts, maintaining social cohesion, etc. This clearly looks differently depending on the social group. Recall that the social group is the raw material for the public machine. Similar to the third end above, the people who take part in the public machine need to do so in a

way that best vindicates our reasons. This just is good functioning. So, it seems that we have all four ends Hursthouse advertised through appeal to the public nature of our reasons.

Concluding Remarks

I have argued that rationality commits us to the publicity of reasons. Further, this publicity of reasons commits us to the social character that we aimed to show. Some may claim that this project is constructivist and not naturalist. This is somewhat correct. I say “somewhat” because it depends on how we define constructivism. While no debate over definitions will be given here, we can note something important about my project. To start, nothing in the vindication process entails that you will form the right reasons. Vindication formed reasons as part of the quest for the right reasons, but we can certainly fail in this quest while still having reasons. If we are looking for right reasons in constructivism, then this project is no such theory. Further, many constructivist views can be understood as accounts of what constitutes correct *reasoning*. My view is clearly a case of this since answering the skeptic and vindicating reasons just is what reasoning is about. This account, however, allows for a naturalist metaethical view depending on external facts that affect the reasoning process.¹⁸⁷ I have only given a skeleton of correct reasoning and, with this, a naturalist view is possible. This reinforces the fact that my project is “somewhat” constructivist.

Bibliography

Annas, Julie. *Intelligent Virtue*. New York: Oxford University Press, 2011.

Baehr, Jason. *The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology*, New York: Oxford University Press, 2011.

Copp, David and David Sobel, “Morality and Virtue: An Assessment of Some Recent Work in Virtue Ethics,” in *Ethics*, Vol. 114, No. 3. University of Chicago Press: 2004. 514-554.

Hursthouse, Rosalind. *On Virtue Ethics*. Oxford: Oxford University Press, 1999.

Korsgaard, Christine. *The Sources of Normativity*. New York: Cambridge University Press, 1996.

------. *The Constitution of Agency*. New York: Oxford University Press, 2008.

Labar, Mark. “Aristotelian Constructivism” in *Social Philosophy and Policy* vol. 25 no. 1 (2008). 182-213.

¹⁸⁷ Southwood (forthcoming).

Lott, Micah. "Why Be a Good Human Being: Natural Goodness, Reason, and the Authority of Human Nature," in *Philosophia* vol. 42 no. 3 (2013).

MacIntyre, Alastair. *Ethics in the Conflicts of Modernity*. Cambridge: Cambridge University Press, 2016.

McDowell, John. "Two Sorts of Naturalism," in *Virtues and Reasons*. Edited by Rosalind Hursthouse, Gavin Lawrence, and Warren Quinn. Oxford: Clarendon Press, 1995.

Mill, John Stuart. *On Liberty*, in *The Collected Works of John Stuart Mill*, vol. XVIII, edited by John M. Robson. Toronto: University of Toronto Press, London: Routledge and Kegan Paul. 1977.

Southwood, Nicholas. "Constructivism about Reasons," in *The Oxford Handbook of Reasons and Normativity*. Oxford University Press, edited by D. Star, forthcoming.

Wallace, R. Jay. "The Publicity of Reasons," in *Philosophical Perspectives* Vol. 23, *Ethics* (2009). 471-497.

Wittgenstein, Ludwig. *Philosophical Investigations*, 3rd ed. Trans. G.E.M. Anscombe, New York: Macmillan Publishing, 1968.

Zagzebski, Linda. *Virtues of the Mind*. Cambridge: Cambridge University Press, 1996.

Everywhere Chimerical

Jorah Dannenberg

In order to see around the intimidating structure that morality has made out of the idea of obligation, we need an account of what obligations are when they are rightly seen as merely one kind of ethical consideration among others. This account will help to lead us away from morality's special notion of moral obligation, and eventually out of the morality system altogether.

(Williams 1985, p. 182)

I. Introduction

In this paper, I take up in earnest a challenge Williams issues in my epigraph, to develop an account of obligation that can serve as an alternative to the more familiar one now dominant in mainstream moral philosophy. Ultimately, I believe such an account can be built around an idea suggested, albeit in quite different ways, by both P.F. Strawson and David Hume. The idea is that our capacity for motivation by moral obligation can be illuminated through serious reflection upon its relation to the way we are moved to respond to the needs and interests of particular others within more intimate and familiar forms of interpersonal interaction. The strategy for understanding moral obligation thus runs counter to the tendency to emphasize the discontinuity between how morality is supposed to move us, and the way we are moved to act for one another by more partial or intimate feelings. It does so, however, not in the manner preferred by some moralists, who pursue continuity by letting morality insinuate itself further into the structure of interpersonal life. The aim is just the opposite: to go as far as possible in articulating the *sui generis* structure of a system of more partial motives, and then construe our capacity to be moved by a sense of moral obligation as an enrichment of the system.

I will make limited progress today on the large task of developing a complete account. I hope to accomplish two preliminary aims, each of which helps to pave the way. The first of these aims is largely negative: I want to characterize in greater detail the mainstream picture of moral obligation I mean to resist. According to the mainstream, the essential feature of any moral obligation is that it constitutes an inescapable constraint on practical reasoning. This claim about moral obligation's essence combines two ideas, both of which are rarely questioned. First, that the basic form of motivation in question must involve heeding a (putative) claim to authority. And second, that what makes *morality* the sort of thing it is – what we might call *the mark of the moral* – is that its power to oblige comes from its special claim to a form of authority that is both absolute and unconditional. Taken together, these ideas imply a familiar picture of how movement by moral obligation happens: an agent deliberating about what

to do recognizes that she is unconditionally required to treat moral considerations as decisive grounds for action, and is moved by that recognition to act accordingly.

The second, more positive aim for today involves beginning to push back against this familiar picture, by following some breadcrumbs Strawson leaves us in *Freedom and Resentment*. There, Strawson suggests there is a way we are moved by feeling bound or obliged to those with whom we are connected by more partial ties, telling us that such motives should be thought of as the agential “correlates” of his famous reactive attitudes. These “correlate” motives take the form of “demands on oneself *for* others” (my emphasis). Strawson’s idea here, once it is suitably cleansed of a tendency to read him through too moralistic a prism, points the way to an understanding of bondage that is quite novel. The emphasis is not on authority, but rather on the way another person’s point of view can, especially in context of love, loyalty, trust, friendship, or affection, simply be internalized within one’s own motivational economy, in such a way that its felt urgency reliably leads one to act for her and her interests. My claim is that this basic sense that one is bound or obliged to act for another can provide an attractive way of thinking about the form of motivation exhibited when someone feels herself *morally obliged*. In other words, the idea is to use this “feeling of bondage” to a particular person to whom one is partial as the blueprint for how a distinctively moral sense of obligation moves us to act for those with whom we otherwise share no special ties.

It is in this way that our capacity for being moved by moral obligation can be regarded as an enrichment of our capacity for the kind of non-moral motive on which it is patterned. Moreover, the relation between these two forms of motivation suggests that such an enrichment is underwritten by a certain *rationale*. The idea here is basically the Humean one, which animates his discussion of the so-called artificial virtues in the treatise: against a backdrop where we are readily moved by forms of partial concern that can lead us to feel bound to act for particular others, we need a moralized sense of being bound or obliged, because it enables us to regard certain action-toward-others with whom we do not otherwise share any special ties as significant, in a way that effectively “mimics” how we regard action done for the sake of those to whom we are partial.

For the purposes of this paper, I can do no more than state this Humean idea about moral obligation’s rationale, though it will be useful to have it in mind in the course of the discussion. Though I will not get around to explicating or defending it, I do hope that by the end of the paper we will at least be in a position see where such a defense would fit. This in itself will be something of an achievement, since it amounts to raising a question about moral obligation that, according to some, it is a mistake to even ask.¹⁸⁸ I will therefore return to these issues about moral obligation’s rationale in the final section, as a way of taking stock.

II. The Mainstream Conception

¹⁸⁸ (Prichard 1912)

The relation between the account I want to advance and the views of the mainstream can be helpfully understood in terms of the familiar distinction between a more or less shared *concept*, and competing *conceptions*.¹⁸⁹ Our concept of a moral obligation is of something that structures and regulates a decent person's thoughts and actions in certain familiar kinds of circumstances. A person's sense of moral obligation reliably leads her to act in ways that we approve of and/or regard as right, often in the face of what would otherwise have been countervailing motives. Sometimes, for instance, a person must honor a difficult promise, respond to a stranger's distress at some cost to herself, or tell an embarrassing truth. What seems essential in situations like these is that a decent person should see herself as in some way bound by having given her word, the urgent needs of another human being, or the value of honesty. A particular *conception* of obligation, then, will provide us with a detailed understanding of just what this sense of oneself as bound comes to. How does it move one to act, when it does? How is being moved in *that* way related to other motives that predominate elsewhere in life? What should we make of the fact that there is this distinctive way of being moved to act in such circumstances, the disposition to which we work hard to cultivate in ourselves, and encourage and admire in others?

The mainstream of moral philosophy has its own familiar answers to these questions. As already indicated, they tend to stem from a view about the nature of morality, as making an absolute and unconditional claim to authority in practical reasoning. The mainstream conception of moral obligation is, in effect, simply what it is supposed to be like for us to have an encounter with morality so-conceived. In deliberating about what to do, an agent recognizes that she is unconditionally required to act in some way, and is moved appropriately by that very recognition. This conception of moral obligation is familiar from Kant's moral philosophy, though he is by no means its only advocate. Indeed, to a great extent, it seems to be the only conception taken seriously by moral philosophers nowadays.

Yet Kant himself offers us little in the way of direct argument for this conception. To the contrary, it is assumed in the statement of the problem that Kant's foundational arguments are meant to address, which aim to establish not *that* morality takes this form, but rather how it *can*.¹⁹⁰ Kant's modern heirs tend to follow him in seeing it as more or

¹⁸⁹As discussed in (Rawls 1971) p. 5

¹⁹⁰ The aim of the *Groundwork*, after all, is to identify and legitimate morality's supreme *principle*, and so presumes it is a principle we ought to be looking for. One might reply that the purpose of the first part of the *Groundwork* is to unpack common-sense moral ideas, in order to show that we are antecedently committed to the idea of a categorical imperative. Kant indeed begins by simply observing that we single out for special praise actions motivated by a distinctive sense of moral purpose, in the face of resistance felt by the agent. This much can be part of a shared *concept* of acting from a sense that one is in some way morally bound – it is, after all, the very thing that struck Hume as remarkable about the morally admirable motives exhibited when, for instance, one restores another's property in secret, or keeps a promise to a distasteful promisee. What can't be part of a shared concept is Kant's claim that what we admire in such circumstances must take the form of obedience, and moreover obedience to some source of commands or imperatives that, if they are to qualify as genuinely moral, must be without condition. "Here mere conformity to law as such, without having as its basis some law determined for certain actions, is what serves the will as its principle, and must so serve it, if duty is not to be everywhere an empty delusion and a chimerical concept." Kant, *Groundwork* 4:402 The passage is striking when contrasted with what Hume says, "What affects us, we conclude, can

less a matter of incontrovertible definition, that morality issues commands, that these commands purport to be categorical, and so that if there is to be anything that counts as a bona fide experience of moral bondage, it can only take the form of rational necessitation in the face of some inescapable principle or norm.¹⁹¹

Where today's non-Kantians tend to demur, however, is with the strategies Kant and his followers pursue to vindicate this conception. Take, for instance, a self-styled (neo-)Aristotelian like Michael Thompson. The "puzzle about justice" that Thompson has recently articulated depends essentially on insisting that morality be unique among systems of norms in how it puts any two human beings "as such" in a special sort of practical relation whereby they will have certain *unconditional* obligations to one another.¹⁹² Thompson's articulation of the puzzle is novel in its emphasis on the "to" – the so called "directedness" or "bi-polar" character that at least many moral obligations have. But in fact, what makes these obligations *puzzling* is that, in all other cases but the moral one, they appear to obtain only when certain metaphysical conditions are met first. At the heart of his puzzle, one thus finds a question undeniably Kantian in its form: How is a system of genuinely *unconditional* "bi-polar" obligations possible?

On the other hand, one can read today's "moral realists" as disagreeing only in insisting that morality is something we need not bother trying to vindicate at all. We are entitled to believe in the unconditionally obliging character of morality, because its existence as part of the furniture of the universe provides the best explanation of how human beings think and behave. The important point here is not about the merits of this sort of "epistemic" argument for the "existence" of morality, but about *what* it is that the realists say we are entitled to take on faith. It is not different than what Kant and his heirs see as in need of a particular kind of vindicatory defense. As Christine Korsgaard rightly notes, the main difference between moral realists and Kantians would appear to be that only the latter are *worried*.¹⁹³

Indeed, so ubiquitous is this conception of moral obligation as the experience of an unconditional claim to absolute authority within practical reasoning, that to be a "Humean" about morality nowadays is, it would seem, to agree with Kant that our experience of morality appears to take this form, and then offer an error theory by appeal to some hypothetical imperative whose condition runs unusually deep.

never be a chimera; and, as our passion is engaged on the one side or the other, we naturally think that the question lies within human comprehension; which, in other cases of this nature, we are apt to entertain some doubt of." Treatise 3.1.1

¹⁹¹ Nagel, for instance, echoes the passage from Kant above when he writes "One must reach a requirement... from which it is not possible to escape by begging off. *It is natural to suppose that principles of this sort underlie ethics, if it exists.*" (Nagel 1970) See also Korsgaard's introductory discussion of "the normative question" (Korsgaard 1996), as well as Chapter 2 of (S. L. Darwall 2006) and essays 1-3 of (S. Darwall 2013).

¹⁹² (Thompson 2004)

¹⁹³ (Korsgaard 1996) p 41-2.

III. Strawson on Resentment, Indignation, and Feeling Bound or Obligated

To be a genuine alternative to this conception of moral obligation and how it moves us, we need a model of what it means to feel *bound* that rejects, or at least plays down, the rationalist assumption at its heart, according to which an experience of bondage must take the form of heeding authority. I believe P.F. Strawson gives us a glimpse of what such an alternative might look like, in some brief but suggestive remarks he makes about our sense of obligation in *Freedom and Resentment*. Here is the passage I have in mind:

the picture is not complete unless we consider also the correlates of [the reactive] attitudes on the part of those on whom the demands are made, on the part of the agents. Just as there are personal and vicarious reactive attitudes associated with demands on others for oneself and demands on others for others, so there are self-reactive attitudes associated with demands on oneself for others. **And here we have to mention such phenomena as feeling bound or obliged (the ‘sense of obligation’);** feeling compunction; feeling guilty or remorseful or at least responsible; and the more complicated phenomenon of shame. (Strawson 1962)

As he reminds us in the first part of the passage, Strawson divides our felt reactions to one another’s motives into two basic kinds. There are the more “invested” feelings that are the article’s main focus: feelings we experience as a constitutive part of more involved, personal relations. Distinct from these, and explained in terms of them, are the more detached, distinctively *moral* reactions, which Strawson characterizes as “vicarious analogues” to the participant feelings. *Resentment* is the leading example of the former kind of attitude, along with feelings like *gratitude*, and *love*. I want to pause on just this point before getting back to the idea of a “correlate” because, despite Strawson’s insistence on the distinction, there is a tendency to read him in a way that elides the difference between the moral and non-moral types of reactive attitude. The following passage from Steve Darwall discussing *resentment* is an illustration of this common sort of conflation.

Resentment... is felt in response to apparent *injustice*, *as if* from the victim’s point of view. We resent what we take to be *violations* against ourselves or those with whom we identify... Resentment seems to be warranted simply by the other’s conduct and by what you and he can *validly*, that is, *authoritatively claim* and demand of one another.” [68, *The Second Personal Standpoint*, my emphasis]

Darwall goes on to say, several pages later

Resentment is felt as if in response to a *violation* of a *legitimate claim* or *expectation*, and not simply as directed toward the *violator*, but as implicitly addressing her.” [72, my emphasis]

These passages appear amidst Darwall’s discussion and elaboration of what he labels “Strawson’s Point” – that the desirability of manifesting a particular reactive attitude on a given occasion is not, as such, part of the warrant for the attitude. To think, for instance, that the question of when gratitude is warranted turns on when feeling it towards a benefactor will do any *good* is to fix on reasons for gratitude “of the wrong kind” (and thus, not reasons *for gratitude* at all). Yet that point can be made without any of the additional, normatively freighty terms Darwall injects injects. Quite clearly, from Strawson’s point of view, Darwall’s characterization of resentment appears unnecessarily moralized.

Consider, as an illustration of genuine (i.e. participant) resentment, its manifestation in the context of romantic rivalry. When I resent the actions of a rival romantic suitor, I need not think he violated any claim I might legitimately have upon him, that he not pursue the object of my affection. What he does is in no way wrong. Indeed, even the weaker thought, that my resentment commits me to somehow thinking he had some reason to act otherwise does not hold: in the case of romantic rivalry, I am if anything committed to the opposite.¹⁹⁴ All that seems essential to make resentment an intelligible response is a disturbance in what might otherwise have been a relation of mutual goodwill between us, which I in some way wanted, hoped for, or would have liked. My resentment just is the registration, in feeling, of such a disturbance.

In light of this, Darwall’s remarks might be thought to provide a better characterization of *indignation*, construed as resentment’s moral analog. But even that would be a stretch, from Strawson’s point of view. For Strawson’s avowed commitment are considerably more Humean. What makes for *moral* feeling is that it is the sort of thing felt by those who contemplate another’s motives in a suitably detached way.¹⁹⁵ As a general matter, our capacity for reasoned judgment does not have any intrinsic authority over this or any other feeling – if anything, it is the opposite. In light of this, we must be very careful what we do and do not think we mean in our use of normative notions like that of *claim*, *right*, *validity*, or *legitimacy*. It is crucial to Strawson’s picture that, though our emotional responses to one another’s motives may acquire a high degree of rational

¹⁹⁴ (Nagel 1970) p. 85 “Resentment... embodies the judgment that another is failing to act on reasons with which one’s own needs provide him.”

¹⁹⁵ “What we have here is, as it were, resentment on behalf of another, where one’s own interest and dignity are not involved; and it is this impersonal or vicarious character of the attitude, added to its others, which entitle it to the qualification ‘moral’.”

structure and articulation, the fuel on which the entire system runs is still, at bottom, a matter of basic human feelings.

Suppose my upstairs neighbor likes to watch his television at all hours at a volume that disturbs me. I politely ask him if he would consider turning it down in the evenings when I am home, but he is unmoved by my request. I will resent the lack of consideration he shows me. Moreover, my resentment in this case *is* likely to be coupled with a more detached reaction of indignation, which may be felt by me or anyone else “vicariously” when viewing the situation impartially. This sort of inconsiderateness between neighbors is something we tend to disapprove of in a distinctively moral way. Yet as I read Strawson, this feeling of moral condemnation is not itself a *product* of any *separate* judgment that I am entitled to better, or have the legitimate authority to expect or demand something he refuses to give. Such judgments are better construed as, as we might put it, embodied in the feeling, and to the extent that one might hope to legitimate them by appeal to conditions of their correctness, one makes reference to nothing further than the way that the overwhelming weight of *normal* human feeling is on my side rather than his.

Central among Strawson’s aims, after all, is turning down the volume on classical debates between libertarian and determinist views of human agency, and their implications for questions of moral responsibility and legal punishment. The original move to the *participant* attitudes is intended precisely as a check on the tendency to see the issues in overly rationalistic way, by showing what it would take to seriously entertain questions about the justification and/or dispensability of our being prone to certain kinds of feelings towards one another in response to the motives we exhibit. It may seem possible, at least in the seminar room, to think there is a “live” question of whether, if it turned out that we are all just a product of our genes, or upbringing, or physical reactions among the atoms in our brains, it could really be correct to punish a criminal or blame a wrongdoer. But Strawson is relying on us to lose our sense that there is any need for legitimation, when focus is shifted to how you will feel about your mother if she forgets your birthday, or your spouse if he takes your turn to walk the dog. The strategy is then to return to the more “sophisticated” questions of *moral* responsibility and desert, but without losing our sense that reason cannot get underneath our commitment to the feelings that fund the entire enterprise.

Thus, as I read him, Strawson’s argument requires eschewing any robust appearance/reality distinction, applied to the normative aspects of our practices of evaluation and response, even when those practices take a distinctively moral turn. Provided we stay within the practice, we can of course accommodate notions of *justice*, *legitimacy*, or *validity*. But it is important to see that *all that lies behind these notions* in the way of a “standard of right” is something like the intelligibility, in the sense of the mere humanity, of the system of feelings in question. Crude appeals to the good consequences of that system of felt reactions are misguided, but so too is any serious attempt to refer them for review to anything outside the system itself.¹⁹⁶ The “rules” for

¹⁹⁶ “...such a question [of justification in terms of an independent standard] could seem real only to one who had utterly failed to grasp the purport of the preceding answer, the fact of our natural human commitment to ordinary inter-personal attitudes. This commitment is part of the general framework of

when one person may *legitimately* insist upon, *validly* expect, or *justly* demand anything in the way of regard from another, whether in a context where morality gets a foothold or not, just *are* a matter of the regularities in these patterns of felt response. The sense in which they are rules is that the normal among us will tend to internalize these patterns, in such a way that they come to guide and structure our reactions. In other words, it is only by way of the system itself that we – which is to say, the *statistically* normal among us – validate, justify, and legitimate, our attitudinal responses towards one another, including both the more invested, personal feelings and the more detached, moral ones.

It is within this system that the feeling that one is bound or obliged should be understood to operate, as a kind of “correlate” to the felt reactions of others. What Strawson should be taken to mean by this, I believe, is that a normal person will often feel moved to act by a kind of internalization of the feelings she anticipates will be directed towards her by another, in accordance with the system so-understood. In other words, the feeling of bondage comes from one’s registering that there is – or will be – a felt significance to one’s performance of an action for another, whether or not one is oneself in any way otherwise moved to do it.

Quite clearly, such a “correlate” can operate, and with considerable felt urgency, even in situations that will not produce any *moral* response in the form of the detached feelings of praise or condemnation. I might feel bound, for instance, to show up at your dinner party, even though I tend to find your other friends tiresome, and this feeling can be a powerful and effective motive. This needn’t (and ordinarily won’t, barring special circumstances e.g. the making of an explicit promise) imply the entire community takes any “vicarious” interest. Rather, my sense of being bound to go is simply the prospective, internalized counterpart of you in particular and your feelings in response to me in particular – your anticipated resentment if don’t go, or gratitude if I do, which will constitute your emotional registration of what my motivation reveals about how I regard you.

Nor should the form of the motive itself be taken to imply any putative claim to authority. In the case of the party, for instance, perhaps I know that you are counting on me to attend, and my sense of the significance of disappointing this expectation is internalized in a way that it will compete with other potential motives, like my desire for a quiet evening alone. Deliberation may be understood in terms of choosing between these motives and, when I do choose to attend, there may be a sense in which my self-interested desire is “overruled” by the feeling that I am bound to you to go. But what is doing the “ruling” here is really me, in selecting between two possible grounds for action, neither of which is inherently more rational or “overriding” as a matter of its form. That I choose the party over solitude because I know what my attendance means to you reveals something about what I regard as important. Having chosen to attend, I may of course articulate my experience in the language of necessitation, as in “I would have preferred to stay home, but really felt I *had* to show up for Mary.” But often, though not

human life, not something that can come up for review as particular cases can come up for review within this general framework.”

disingenuous, there will be a fair amount of exaggeration in such talk, insofar as I of course literally could have chosen otherwise, and sometimes do. On the other hand, even when the language of necessitation is not an exaggeration, it need not take the form of distinctively *rational* necessitation. If I feel I absolutely *must* give Mary my kidney, this may be literally true, without anything (other than my own sense of what I can and cannot live with) demanding it.¹⁹⁷

With this picture of how we might feel bound or obliged to a particular other, in conjunction with Strawson's notion of the mark of the moral as a matter of vicariousness of feeling, a picture of moral obligation begins to emerge. What constitutes a sense that one is distinctively *morally* bound or obliged will depend on the internalization of not just any *particular* person's feelings, but rather of the kind of detached response that itself mirrors those of the offended or benefited. The responses, that is, which are manifested in reactions of disapproval and indignation felt by others in general, "on the offended party's behalf," can be correlated on the part of agents, in much the way the more particular attitudes like resentment or gratitude are. A person who feels obliged, say, to tell an unpleasant truth to a stranger (I am the one who scratched your car), or who declines to exploit another's mistake in a commercial transaction (I think you've given the wrong change), acts from internalized correlates of the sort of condemnation or praise that any of us is apt to feel upon the detached contemplation of the actions she undertakes or foreswears. As with the interpersonal case, the urgency with which these motives will be felt will be a matter of how strongly the agent has internalized the point of view from which these distinctively *moral* feelings are experienced. The moral quality of the feeling comes not from it making any special claim to authority: certainly not any unconditional one. As with the original sentimentalists, a person's conscience might here be represented not as a voice she obeys, but rather as view that sees her, and whose approval she wants to earn. When there is conflict between the deliverances of conscience and other motives, the degree to which a person is apt to let morality "overrule" other potential sources of motivation will be a matter of the broader composition of her character: of who she is, and what she can and cannot live with.

V.

These ideas fall short of a complete account of obligation, but I do hope to have done two things in this paper. First, I have said what the defining features are of the mainstream conception of obligation and how it moves us. Second, I have urged that Strawson, read carefully, points the way to a genuine alternative to that mainstream conception. We find in Strawson a way of understanding what it means to feel oneself morally bound to do something, in way that is both recognizably *moral* and recognizably a form of *bondage*, yet which does not take the mainstream form of an inescapable constraint on practical reasoning.

¹⁹⁷ See the essay *Practical Necessity* (Williams 1981)

As I said at the outset, I want to close by returning to the question of moral obligation's *rationale*. What I mean by this is not an answer to the question of what *justifies* us in regarding ourselves *morally* bound to do certain things; rather, I mean to raise a question about moral obligations point, or purpose. What, in other words, is our capacity for movement by a sense of moral obligation *for*?

As I noted above, when morality's power to bind is viewed as a matter of its unconditional authority, some will maintain this question is confused. Strawson himself says some rather confounding things on this. On the one hand, he seems to suggest that the idea that what I am trying to get at by the question of the *rationale* may not be altogether a live one, by which I think he means there is no real possibility of exiting the system. On the other hand, he says, the system of attitudes and motives in question are "humanly necessary." While that statement (rightly, I think) may be taken to rebuff the request to *justify* the system, it might still leave it open to us to ask: necessary for *what*? It is at this point that one begins to see a place, at least when it comes to the capacity for motivation by our obligations, for a defense of Hume's answer, as laid out in his discussion of the moral motives that in the *Treatise* he labels "artificial."¹⁹⁸ The necessity of our capacity for being moved by moral obligation shows itself in how that capacity enables us to live together with others, with whom we do not otherwise share special bonds of feeling. It does this, however, not in the manner implied by the mainstream view: by constraining our unruly or unreliable natures. Rather, by internalizing a healthy dose of the perspective from which moral reactions are felt, and coming to be reliably moved by that perspective, we are able to see and treat others as such *as if* they were our family, or lovers, or friends.

Let me close by returning to Williams, who no doubt agreed that we could, and should, make sense of the question of moral obligation's rationale. But he seems to have held, like Nietzsche before him, that once we find ourselves in a position where the question can be asked, there is little hope of giving an answer that is not corrosive. This is because he thought that any honest appraisal of the concept's rationale reveals it as one that serves only shameful impulses: the drive to vent our inhumane instincts, to control others, or to protect ourselves from misfortune. Williams therefore would have balked at the idea of developing an account of obligation and how it moves us that retained any particularly *moral* sense. His point was clearly not just about terminology: having understood what moral obligation is *really* for, we ought to (try to) do away entirely with the particular – he would have added "peculiar" – system of ideas within which any distinctively moral sense of obligation is at home.

In case it is not obvious, I am deeply sympathetic to these strains of argument. Yet while I agree that mainstream ideas about moral obligation have had, and continue to have, a pernicious influence both inside and outside philosophy, I am not (yet) prepared to cede so much ground to our adversaries. I am, in truth, agnostic. I wonder whether even those of us who are cynics about morality in its traditional form(s) may do better to

¹⁹⁸ To redeploy a bit of Hume's own terminology, we might for instance say that a person who acts from her sense that she is obliged e.g. to honor a promise she has made exhibits a motive that is the "artificial" simulacrum of the motives whereby we display loyalty for those to whom we are more "naturally" partial.

fight for the reformation of our ideas about moral obligation, rather than relinquishing the concept altogether.

This is not to concede that my proposal is entirely revisionary – it does not amount to an attempt at radical redefinition, though it will perhaps be felt that way among moral philosophers.¹⁹⁹ “Our” system of moral ideas is a messy one, with many more and less dominant currents and themes. We therefore must sometimes choose where to focus our attention as theorists, and how to let ourselves be guided as agents. One thing worth trying, even for those of us who find ourselves in the dissenting minority, is to seize upon one kind of undercurrent within that messy system that is often ignored, and to refine and champion it, with the hope that doing so might make the choice of whether, or how, to live with *some* notion moral obligation a clearer one.

References

Darwall, S. (2013). *Morality, Authority, and Law: Essays in Second-Personal Ethics I*: Oxford University Press.

Darwall, S. L. (2006). *The Second-Person Standpoint: Morality, Respect, and Accountability*: Harvard University Press.

Hume, D. (1978). *A treatise of Human Nature*. Oxford ;New York: Clarendon Press.

Kant, I. (2011). *Immanuel Kant: Groundwork of the Metaphysics of Morals: A German--English Edition*: Cambridge University Press.

Korsgaard, C. (1996). *The Sources of Normativity*. Cambridge: Cambridge UP.

Nagel, T. (1970). *The Possibility of Altruism*: Princeton UP.

Prichard, H. A. (1912). Does Moral Philosophy Rest on a Mistake?. *Mind*, 21(81), 21-37.

Rawls, J. (1971). *A Theory of Justice*: Harvard University Press.

Stevenson, C. L. (1938). Persuasive Definitions. *Mind*, 47(187), 331-350.

Strawson, P. F. (1962). Freedom and Resentment. *Proceedings of the British Academy*, 48, 1-25.

Thompson, M. (2004). What is it to Wrong Someone? A Puzzle About Justice. In *Reason and Value: Theme From the Moral Philosophy of Joseph Raz*.

¹⁹⁹ Of the sort suggested by (Stevenson 1938), for instance.

Williams, B. A. O. (1985). *Ethics and the Limits of Philosophy*: Harvard University Press.

Williams, B. A. O. (1981). *Moral Luck: Philosophical Papers, 1973-1980*: Cambridge University Press.

SOLIDARITY AND COMMUNITY

Mike Zhao

Introduction

A prisoner-of-war is secretly offered early release by his captors, who know that his father is an important figure in the military. He decides to remain in captivity in solidarity with his platoon mates.

A young girl is undergoing chemotherapy for leukemia, which has caused her hair to fall out. Her parents and older siblings shave their own heads in solidarity with her.

Workers at an equipment factory go on strike to protest an upcoming round of layoffs. Workers at another factory in the same town, whose jobs are not in danger, go on strike in solidarity with them.

Marie is a young woman living in East Germany in the late 1980s. On the night of November 9, 1989, the Berlin Wall opens, reuniting East and West Germany after 44 years of separation. A group of young people from both countries have climbed atop the wall and spontaneously begin singing the *Deutschlandlied*, a song of German unity. Marie, watching from her own home, sings in solidarity with them.

Solidarity is at once widespread and puzzling. The examples I've just given show that it's a ubiquitous human phenomenon: the idea of "being in it together" with others, of suffering what they are suffering or rejoicing in what they are rejoicing in, is instantly intelligible, and seems to have its roots somewhere deep in the human psyche. But questions arise once we start to think about solidarity. First, why do people act in solidarity with others? Acts of solidarity *can* be directed toward social change, but they need not be. They *can* have a function of expressing unity with a group, but they need not. After all, expressing a mental state requires intending that others believe that one has it; but as some of these examples show, one can act in solidarity with others without even intending that they know.

Furthermore, acts of solidarity often consist in *doing* or *undergoing* what those that one acts in solidarity with are doing or undergoing: the prisoner-of-war acts in solidarity with his platoon mates by continuing to undergo captivity together with them; Marie acts in solidarity with other Germans by singing with them. The question arises of what prompts us to act together with others in cases like this, and why these acts qualify as solidaristic ones.

Finally, many of us find it admirable in many cases when people act in solidarity with others. We might think that the prisoner-of-war and the young girl's family are exhibiting *some* positive character traits in acting in solidarity, that they're responding correctly to

some value. And this raises the question of *which* virtues they're exhibiting, *which* values their actions are responses to. Of course, I take it that many might have instead a neutral or even negative reaction to many cases of solidarity. We might think that the agents are doing something narcissistic or otherwise self-regarding, trying to impress others or themselves with their own virtue. Nonetheless, many of us think that many cases of solidarity—at least those not motivated by self-regarding concerns—are admirable, and we need an explanation for *why* we think this. And even if we think many acts of solidarity are simply attempts at virtue-signaling, that requires that we think (or at least think that others think) that there are more-or-less *genuine* acts of solidarity that do exhibit some virtue, which the self-regarding cases are piggybacking on.

In this paper, I want to offer an account of solidarity, which will answer these questions: what prompts acts of solidarity, why acting together often counts as acting in solidarity, and what values (if any) one responds to in acting in solidarity. One consequence of what I have to say is that solidarity is not distinctive, but one instance of a more general relationship of *community* that people can stand in to each other.

This paper has three sections: in §1, I start by looking at a subclass of solidaristic acts, which I'll call *self-depriving* acts of solidarity; I'll discuss and criticize a recent treatment of these acts; in §2, I provide my own diagnosis of these acts; and in §3, I use this diagnosis to provide a more general account of solidarity.

1. Self-deprivation

I want to start by drawing attention to one feature present in many acts of solidarity, including some of the examples I gave above. Consider the prisoner-of-war's acting in solidarity with his platoon mates by remaining in captivity, or the young girl's family acting in solidarity with her by shaving their own heads. Both of these acts consist in the agents' voluntarily depriving themselves of something that is available to them (freedom, having hair), but that the objects of their solidarity (the other members of the platoon, the young girl) are deprived of. I'll call acts like this *self-depriving* acts of solidarity.

Of course, not all acts of solidarity involve self-deprivation. The workers engaging in the solidarity strike aren't depriving themselves of anything that the workers at the other factory are deprived of. And in the Berlin Wall example, Marie certainly isn't depriving herself of anything that other Germans are being deprived of by singing. Nonetheless, I want to focus first on self-depriving acts of solidarity because they form a particularly interesting class of solidaristic acts.

Let me give two more examples of self-depriving acts of solidarity:

A gay couple living in the US have the legal right to marry, but decide not to marry in solidarity with gays around the world who do not have the legal right.

In an episode of the TV show *Black Mirror* set in the 2040s, humanity has the technology to upload the consciousness of the dying into a virtual reality, thereby prolonging life indefinitely. A woman recounts that her daughter died before the technology became available, and that her husband passed up the offer of technologically-aided immortality because their daughter did not get the option, choosing to die naturally in solidarity with her.

These examples draw attention to a few points. First, acts of solidarity need not be public, directed toward the object of solidarity or some third entity. Second, solidarity need not presuppose any goal or interest that the agent and object of solidarity share, which the act of solidarity advances. What exactly, for example, is the goal shared by the man and his dead daughter, which is furthered by his choice to die naturally? In fact, acts of solidarity can often have an *anti*-consequentialist or leveling effect, making the agent worse off without making the object of solidarity (or anyone else) better off. The man's daughter, for example, is not better off now that her father has decided to die naturally; it's not as if she's comforted by his decision, since she's not even alive to know about it. A third point is that there need not be anything both the agent and object of solidarity are standing together *against*, as is sometimes presupposed: what exactly are the father and daughter jointly opposing? Fourth, the objects of solidarity need not be victims of injustice or wrongdoing: perhaps the platoon is being held unjustly, and perhaps gays in other countries are being wrongly denied their right to marry, but the woman who died before immortality became technologically feasible is not necessarily the victim of any wrongdoing.

I take it that many readers will form a positive assessment of the agents of solidarity; you might think that they are acting virtuously. Nonetheless, their behavior might seem rationally inexplicable, given the usual moral theories. Consequentialism seems to recommend against self-depriving acts of solidarity, since (as we've mentioned) they often make things worse for the agent without making them better for anyone else. Similarly, it seems hard to come up with a deontological theory that could rationalize these acts. I might have various positive and negative obligations toward my children: not to harm them, to love and provide for them, and so on; but to deprive myself of something that they are deprived of does not seem to fall in the scope of any of those obligations.

Nicolas Bommarito has recently given an insightful discussion of these kinds of cases.²⁰⁰ Bommarito's account of why such acts are virtuous is that they demonstrate *concern* on the agent's part for the object's wellbeing: the prisoner of war shows his concern for the plight of his platoon mates by deciding to forgo early release, and the gay couple show their concern for gays in other countries by deciding not to get married.

This account strikes me, however, as implausible. Typically, if you act out of concern for someone else, then you try to make that person better off. Again, however, in these cases the acts of solidarity make the object of solidarity no better off, but rather make the agent *worse off*. The man's dead daughter doesn't benefit from his decision, nor do gays in

²⁰⁰(2015), "Private Solidarity," *Ethical Theory and Moral Practice*.

other countries benefit from the American couple's decision; furthermore, the agents know in advance that what they do won't affect the others, so it is strange to think of their decision as prompted by concern for the well-being of other people.

Perhaps one could say something in defense of the "concern" account. One might object, for example, that concern for another's wellbeing need not express itself in wanting to make the object of concern better off; it might manifest itself in lots of different ways. If I hear about a terrorist attack in London, and know that my friend lives in London, I might text her out of concern, just to make sure that she's okay. My motive for *that* consists simply in a desire to make sure that nothing bad has happened to my friend, and does not involve a desire to make her better off. I acknowledge this point, but in response, it still seems like a condition on concern that one cannot have concern for those who (one believes) no longer exist. I might do things to commemorate my deceased grandfather or to further his legacy, but it doesn't make sense for me to do things out of *concern* for him, in his non-existent condition. And if this is so, the man cannot be acting out of concern for his daughter by acting in solidarity with her.

Or, as a second response, one might concede that acts of solidarity really don't have value when neither the objects of solidarity, nor anyone else, is aware of them. After all, one might think that acts of solidarity are supposed to have an *expressive* function, and it's precisely this that is missing when you engage in private acts of solidarity: you can't be *expressing* any attitude of togetherness if you don't intend that others come to believe that you have that attitude. Nonetheless, the acts *do* have value when the object of solidarity is aware of them, since in that case they *do* make the object better off, by comforting her and letting her know that others are supporting her. The young girl with leukemia, after all, receives comfort and strength from her family's public display of support for her.

But in response to this thought, note that it pushes the question a step back: given that we are comforted to learn that others are suffering with us, the question arises of *why* we are comforted. After all, surely it's not just *Schadenfreude*: we don't just enjoy watching people suffer. Rather, we're comforted because the agents seem to be responding to some kind of value. This prompts the question of *what* value that could be, and raises the possibility that the same could be going on in cases in which the objects of solidarity aren't aware that anyone is acting in solidarity with them.

It does seem true that, in the cases of solidarity that we discussed, the agents *do* have concern for the objects of their solidarity. My point is simply that this concern cannot be the *motive* for their actions.

2. Fairness

I want to propose another account of self-depriving acts of solidarity, and I want to start by looking at two features that the cases of solidarity that we discussed above share. What I

say about these cases won't generalize to all cases of solidarity; nonetheless, it will point us to something more general whose presence explains both these cases and other cases of solidarity.

First, both the agent and object of solidarity must belong to a *group* that the agent identifies strongly with: a family, a platoon, or a marginalized group. It makes sense for the young girl's family's to shave their heads in solidarity with her, but it'd be strange if they shaved it in solidarity with another girl, who doesn't stand in any special relation to them. Similarly, it makes sense for the prisoner-of-war to forgo early release in solidarity with his platoon mates, but could we make sense of his doing that in solidarity with prisoners in another platoon, or serving another country? Acts of solidarity require, as a background thought, that the agent and object both belong to some group.

Second, let's think about what the agent's motive is in the examples we mentioned above. Take the case of the man who forgoes immortality because his daughter never got the option. Here, it seems natural to supply him with the following internal monologue: "My daughter didn't get the option to live forever. I don't deserve better than she does, so because she didn't get to live forever, I won't choose to either." Or think of the case of the prisoner-of-war who forgoes early release: "We're all members of the same platoon, and I don't deserve better than they do. So if my platoon mates don't get to be released, then I won't accept release either." Think of general expressions of this thought: "If they don't get to have it, then I won't have it either," or, "If not all of us can have it, then none of us will."

In these cases, concern for the object of solidarity does not play a direct role, as we've already discussed. Rather, my suggestion is that, because the agent identifies with a group that both he and the objects of solidarity belong to, the agent has the thought that *he is not more deserving* (perhaps of something in particular) than the others. This means that if the others are deprived of something, it would be unfair for him to have that thing. In other words, he thinks that if no one else gets their deserts, then it would be unfair for him to get his. This expresses a sense of *humility* and a concern for *fairness*: it expresses humility because the agent takes himself to be no more deserving than others; and it expresses a concern for fairness because, in acting out of solidarity, the agent relinquishes what he thinks it would be unfair for him to have, given what others get to have. Of course, to the extent that concern for other members of the group is often the result of strong identification with the group, concern for the object of solidarity is often present in acts of solidarity, but is just epiphenomenal. Let me spell out the principle of fairness that is operative in these cases of solidarity. I contend that, in these cases, the agent chooses to deprive himself of what others are deprived of because he thinks that *it is unfair for a member of a group to have something if (enough) other members of the group do not get to have it*. So, in particular, it is unfair for *me* to have something if other members of the group do not get to have it.

We need to circumscribe this principle in two ways. First, such a principle applies only in the case of *certain* groups. After all, there are many groups membership in which is morally irrelevant. Take, for example, the group comprising people whose first name starts with the same letter as mine; I am a member of this group. But such a group obviously does

not seem to be the right kind in which acts of solidarity can arise, as the relation between the members of the group is too thin. (I don't give things up just because some Matt somewhere doesn't get to have them.) Nor will just any morally significant group, one that is capable of grounding moral obligations, do. If a buyer of my goods and I sign a contract, then both of us have a moral obligation to uphold the contract; we form a morally significant group. But this doesn't seem like the right kind of relationship that acts of solidarity spring from; I don't really care about what happens to the buyer outside of the very limited scope established by the contract. So the group has to be specified more finely. I'm going to stay neutral on the question of *which* groups solidarity can arise in; we have an intuitive grasp of the answer, and that will do for our purposes.

Second, we need to limit the kinds of *holdings* in the scope of the principle, the kinds of things that the agent thinks that members of the group deserve equally. The fact that I am a member of the right kind of group as others does not necessarily mean that it would be unfair for me to have *anything* that the others are deprived of. Take the case of the gay couple again: that couple thinks it would be unfair for them to marry when many gays around the world cannot. But they may not similarly think that it would be unfair for them to dine at a nice restaurant and see a show afterward simply because many gays in less wealthy countries cannot afford that luxury.

One might think that which holdings the principle covers should relate to what defines the group in the first place. So in the case of groups defined by sexual orientation, one might expect that the principle covers only holdings having to do with sexual orientation: a member of that group might think that it's unfair that his sexual orientation is legally recognized when that of others in the group is not. Similarly, in the case of citizens of the same country, perhaps the principle covers only legal rights, and other things that flow directly from the basic structure of society. So perhaps a citizen of a country might think that it is unfair if he is treated more leniently by the legal system than other citizens are, but not if he is better looking than other citizens.

The restriction seems less plausible when we turn to other groups, however, such as those of friendship or family; in those cases, some kinds of holdings that the principle covers do not seem to be related at all to the criteria for membership in the group. What two friends share is just liking for and some sense of closeness with each other; but people might forgo many kinds of things that their friends are deprived of. Similarly, in the case of a family, the members share ties of marriage and biology. Now, think about the case involving the young girl made bald by chemotherapy. How does the nature of the relationship between family members imply, for example, that having hair is within the scope of the principle of fairness that applies to them? So while that principle must be restricted to apply only to *certain* holdings, it seems that there is no clear way to specify which holdings the claim applies to in terms of the kind of group involved. Again, however, this is fine for our purposes; I just want to note that the principle has to be circumscribed in *some* way.

Now, one way in which this principle is interesting is that it conflicts with a widespread conception of justice, that justice consists in each person's getting what he deserves. Recall what Mill said about justice:

It is universally considered just that each person should obtain that (whether good or evil) which he *deserves*; and unjust that he should obtain a good, or be made to undergo an evil, which he does not deserve. This is, perhaps, the clearest and most emphatic form in which the idea of justice is conceived by the general mind.²⁰¹

And even if one doesn't accept a desert-based conception of justice to the exclusion of other conceptions, one might at least think that it *contributes* to the justice of a state-of-affairs that everyone gets what he deserves. If what I have said is correct, however, then it might be *unfair* for one person to get his deserts, so long as no one else gets theirs.

Before going on, I want briefly to mention one application of the principle of fairness that I posited. One thing that the account helps to explain is the phenomenon of *survivor guilt*, which often occurs after someone has survived something that many of those who went through what she did did not survive: a plane crash, an epidemic, a mass shooting, or an atrocity. There are different possible causes of survivor guilt: not having done all in one's power to help others survive, or having actively hindered others from surviving in order to survive oneself (for example, trampling over others in a mad rush to flee from danger). Both of these cases of survivor guilt can be explained by the subject's being partly causally responsible for others' deaths. But even when these two factors are absent—even when one could not have helped others survive, and did not hinder them from surviving—one can have survivor guilt merely from the thought that *I survived while others died*, and this can seem rationally inexplicable.

I think that we can explain such cases of survivor guilt in terms of the agent's having the thought that it was unfair for her to survive when others died, and that we can explain her having this thought in terms of her accepting the principle of fairness that we appealed to earlier. One piece of evidence for this is that survivors themselves often say that it was unfair for them to survive when others died. As another piece of evidence for this, they sometimes appeal to the thought that they did not *deserve* to survive any more than others did. A woman who narrowly missed getting on one of the planes that crashed on 9/11 said, in an interview, "Why me—out of all those wonderful people? What have I done? I'm not a saint or angel."²⁰² The implicature here is that if she somehow *were* more deserving of survival than others (say, by being a saint), her reaction of survivor guilt would not be as fitting. My account would explain this via the connection between fairness and desert: if she deserved survival more than those around her did, then this would lessen any unfairness of her surviving while others died.

Of course, in explaining these cases of survivor guilt, I'm not saying that those with survivor guilt didn't *deserve* to survive; in fact, it's precisely because *everyone* deserved to survive and many of them didn't that survivor guilt arises in the first place.

²⁰¹(1863), *Utilitarianism*, ch. 5.

²⁰²"Small misses, saved lives: Near misses of 9/11," *CNN*, September 5, 2011.

Communal sharing

To sum up the discussion in the last section, we noticed that in self-depriving acts of solidarity, the agent accepts a principle of fairness that she takes to apply to some group that both she and the object of solidarity are part of and some holding that she has and that the object is deprived of; as a result, she thinks that it is unfair for her to have what the others are deprived of, and voluntarily deprives herself of that thing.

Now, this doesn't account for all cases of solidarity, and I don't intend it to; after all, there are many *non*-self-depriving acts of solidarity. Think about the young German woman who sings in solidarity with other Germans when the Berlin Wall falls. There's no self-deprivation going on in this case, and the fairness account doesn't seem at all to explain why *she's* doing what she's doing. Rather, what seems to motivate the young woman is a sense of unity with a larger entity, one that suppresses the sense of being a separate individual, and so motivates her to do whatever the group as a whole is doing. So the question now is whether we can give a unified account of acts of solidarity, which explains both the self-depriving and these diverse non-self-depriving cases.

I think we can, and I think we can do it by appealing to something more general than acceptance of the principle of fairness, the tendency to treat what happens to other group members in a first-personal way, and the tendency to act together with the group. Perhaps *a certain way of relating to others* in a group generates, in different contexts, *each* of these mental states. If this is so, then using that kind of scheme in interactions with others will explain both cases of solidarity that involve self-deprivation and other cases, and we can give a unified account of acts of solidarity.

I want to start by drawing on something that the anthropologist Alan Fiske talked about. Fiske proposed the idea that people have several *basic mental models* for interacting with other people.²⁰³ We can think of these models as *concepts*, or schemata, for basic types of relationships. Now, thinking of something using a concept guides how we behave toward that thing; so thinking of my father as my father means that I'll behave in certain ways toward him, like not calling him by his first name. Similarly, using one of these mental models guides how I behave toward the other people involved.

The model that I want to focus on is what Fiske calls the *communal sharing* model, which represents members of the group as *equivalent* and *undifferentiated*. By "equivalent," Fiske means that members treat one another as being alike on some important dimension, and overlook the ways in which they are unlike; this leads them to treat one another as, in certain respects, interchangeable. In particular, they react to what happens to others in a first-personal way, as if those things were happening to them; and they take the

²⁰³(1992), "The Four Elementary Forms of Sociality," *Psychological Review*.

members to have equal moral status. And by “undifferentiated,” Fiske means that members experience the group as a cohesive whole, which can subsume their own individual agency.

Fiske takes the communal sharing model to be the most basic of the models, in several senses: children develop and externalize this model the earliest, at infancy; its use is prevalent in the earliest forms of society; and its use corresponds to, according to Fiske, the deepest human needs. (By the way, the other three models that Fiske posits are hierarchy, reciprocity, and fair exchange.)

The term “communal sharing” is something of a misnomer. Fiske posited the model to explain the widespread phenomenon of free sharing of resources within certain kinds of groups, but the model clearly explains much more than that: celebrating others’ joys and mourning their sorrows, feeling pride or shame as a result of what they do, notions of collective responsibility, spontaneous collective action, and so on.

Let me just give an example to make this discussion seem less exotic. Think about the interactions between the members of an immediate family. First, members of a family typically think of one another as having equal moral status: they think that it’s unfair if one member of the family enjoys a significantly higher standard of living, or if the parents treat one child’s wellbeing as more important than the others’. Second, they often have vicarious first-personal reactions toward the actions or experiences of other family members: they feel pride at one another’s accomplishments, shame at one another’s failings, and so on. Finally, there’s often an emphasis on doing activities as a unit. In Fiske’s terms, the members of a family often use the communal sharing model in interacting with each other.

Although Fiske does not say so explicitly, it seems that there are *triggers* for the use of each of the mental models on a given occasion. In the case of communal sharing, that trigger is the awareness of some important feature shared by the members of a group: blood, ethnicity, citizenship, profession, etc. The recognition of sameness along some important dimension then creates the disposition to treat one another as equivalent. These mental models are flexible, and a person can adopt different models on different occasions, or even different models in the course of the same relationship. Thus the members of a family might use the communal sharing model for certain interactions, and a hierarchical model for others: we’ve just said that family members tend to treat one another as morally equal (which indicates use of the communal sharing model), but the parents take themselves to have certain prerogatives and directional obligations (which indicates use of the hierarchical model).

Let me note that beyond the mental models themselves, we can also talk about *groups* in which the use of the models forms a common mode of interaction. As I mentioned, families often think of themselves in communal sharing terms; the military is another example of a group that many think of in communal terms, although (like a family) it has hierarchical elements too: dress is uniform, to downplay the differences between individuals, and military cemeteries consist of lattices of uniform headstones. Of course, complete communal sharing might be an ideal, not fully realized by any actually existing social group; in practice, it will always be limited by the use of competing models, not to

mention ordinary human selfishness. But many groups do exhibit widespread use of communal sharing.

Beyond descriptive facts about prevalent modes of interaction in different groups, there are also *normative* facts about when the use of a model is appropriate. It would be inappropriate on many occasions, for example, for the members of a family to use the fair-exchange model with each other: imagine parents who, before serving their children dinner, ask “What do I get in return?” Similarly, it would be inappropriate for friends interacting as such to use the hierarchical model: for one friend to order the other friend around, for example.

Let us call a group in which communal sharing forms an appropriate mode of social interaction a *community*. When Rawls writes, “In justice as fairness men agree to share one another’s fate,”²⁰⁴ he is drawing on an impulse to think of one’s society as what I have called a community: we are all members of the same society, so let us agree to enjoy the same outcomes. I’ve thus defined “community” to be a partly normative term: it must be *appropriate* for members of a group to treat one another as equivalent and undifferentiated in order for the group to qualify as a community. More colloquially: a community is a group in which members *ought* to treat each other as equivalent and undifferentiated, at least in certain respects.

Although I am using “community” as a term of art, my definition meshes with how other thinkers have defined the term. Ferdinand Tönnies, for example, defines his term *Gemeinschaft* (usually translated as “community”) as a group defined by features shared among the members, and in which there is a sense of identification with the whole. Similarly, Michael Sandel talks about a “sense of community” that members of a society might have, which requires taking membership in that society to form an important part of one’s practical identity.²⁰⁵ Identifying with the society can then serve as the basis for believing that one has certain entitlements (“as Americans, we’re all entitled to certain political rights”) and as the basis for certain reactions, such as national pride or shame.

How does this relate to solidarity? Well, I want to propose that use of the communal sharing model—that is, thinking of a group as a community, one in which members ought to be treated as equivalent and undifferentiated—forms the deepest level of an explanation of acts of solidarity. As I argued in the last section, acceptance of a principle of fairness explains the self-depriving acts of solidarity that we encountered. But because thinking of a group as a community entails thinking of its members as having equal moral status, it also explains why an agent thinks that it would be unfair for him to have things that others are deprived of. So the family of the young girl undergoing chemotherapy think of themselves as, in certain way, having equal moral status as her, thereby thinking that it would be unfair for them to have hair if she doesn’t get to have any.

²⁰⁴(1971), *A Theory of Justice*, p. 102.

²⁰⁵(1982), *Liberalism and the Limits of Justice*.

How does thinking of a group as a community explain non–self-depriving cases of solidarity, though? Let’s think about the cases that we came across. Think first about the solidarity strike case, in which workers at a second factory strike in solidarity with those striking at the first. In this case, what motivates workers at the second factory to strike is a desire to further the interests of the workers at the first factory. But the tendency to treat the interests of others as one’s own also arises from seeing the group as equivalent, which is a part of thinking of the group as a community. Finally, take the Berlin Wall example, in which Marie sings in solidarity with other Germans. Here, what does the work is the thought of the group as undifferentiated: feeling “at one” with the group, which motivates the individual to do what the others are doing.

In sum, thinking of a group as a community explains both the self-depriving and the other acts of solidarity, and we have a unified account to give. Now, one might think that this condition is too restrictive. After all, ordinary use of the term “solidarity” seems to recognize the possibility of being in solidarity with people that one has little in common with: so citizens of the US seem to be able to stand in solidarity with undocumented immigrants, or refugees.

There is also a phenomenon that is at least related, which we might call expressions of *empathetic identification*. These are expressions of identification with a group that one does not belong to, seemingly for the purpose of signaling solidarity with that group. So think about JFK’s declaration, “*Ich bin ein Berliner*,” or the *Je suis Charlie* hashtag that became popular after the terrorist attack on *Charlie Hebdo*. Now, if, as I have said, one can be in solidarity only with those that one takes to be members of the same community, what are we to make of these declarations, if one is not a Berliner or member of *Charlie*? My response is that these *can* be cases of solidarity, so long as the agent thinks that there is a *larger* community that encompasses both herself and the objects of her solidarity. So an American citizen can stand in solidarity with non-citizen immigrants to the country if Americans (on some criterion less stringent than citizenship) count as a community; the thought might be that those who are living in the country deserve to be treated equally. Similarly, if one thinks of humanity as a whole as forming a community, then one can be in solidarity with undocumented immigrants, refugees, Berliners, *Charlie Hebdo*, etc. on the basis of shared membership in humanity.

Conclusion

Let me wrap up. We started out by looking at a class of acts of solidarity, those that involve self-deprivation on the agent’s part. I argued that what motivates these acts is humility and a concern for fairness within a group that the agent identifies with: in light of belonging to the same group, the agent thinks that it would be unfair for him to have what others are deprived of. I used this to build a more general account of acts of solidarity, which explains not only self-depriving, but also other acts of solidarity. On this account, what motivates the agents is thinking of the group as what I’ve called a community.

What's Wrong with Gentrification? A Structural Republican Approach

Rafeeq Hasan

Introduction

Between 2000 and 2003, the number of low-income households in the Bay Area increased by 10%, but the region lost 50% of units defined as affordable for this population.²⁰⁶ The San Francisco Tenants Union estimates that no-fault evictions displaced nearly 1,400 renters in 2013.²⁰⁷ Since 2007, at least 172,000 apartments in New York City have been removed from rent control and placed on the deregulated market.²⁰⁸ The disappearance of affordable housing in urban centers has been described by humanitarian organizations as a “complex emergency” that has led to large-scale “displacement of populations” from their homes.²⁰⁹

Both scholars and political actors often attribute the rising costs of housing, and the attendant movement of lower and middle-class residents away from urban centers, to the process of gentrification. The term “gentrification” comes from the British sociologist Ruth Glass, who in the 1960s observed an influx of people of more affluent and educated backgrounds than their working-class neighbors—‘gentry’—buying properties in inner London that had fallen into neglect and disrepair.²¹⁰ Contemporary geographers treat gentrification as “a generalized middle-class restructuring of place, encompassing the entire transformation from low-status neighborhoods to upper-middle-class playgrounds.”²¹¹

Does gentrification represent a failure of *justice*—that is, a failure of the state to provide for the essential needs and freedoms of all its citizens—and, if so, why? Do citizens of urban areas in relatively prosperous societies have a *right* to urban housing, even if market conditions place the cost of such housing out of reach? Assuming that states ought to ensure that all citizens have *some* housing within proximity to work and family, and that no one remains homeless, do the very same considerations that justify those rights also justify a right to remain in the neighborhoods in which one already

²⁰⁶ <https://www.theguardian.com/technology/2016/jul/07/silicon-valley-largest-eviction-rent-controlled-tenants-income-inequality>

²⁰⁷ <https://www.theguardian.com/world/2014/feb/23/is-san-francisco-losing-its-soul>

²⁰⁸ <http://www.nybooks.com/articles/2017/08/17/tenants-under-siege-inside-new-york-city-housing-crisis/>

²⁰⁹ <http://www.nybooks.com/articles/2017/08/17/tenants-under-siege-inside-new-york-city-housing-crisis/>

²¹⁰ Kate Shaw, “Gentrification: What It is, Why It Is, and What Can Be Done About It,” *Geography Compass*, 2/5 (2008): 1697-1728.

²¹¹ Shaw: 1698.

lives—what has sometimes been called “the right to stay put,”²¹² or “the right to the city?”²¹³

Liberal democracies are committed to private markets in housing. So remedying the phenomena described above would necessitate various forms of restriction on landlords (e.g., rent-control and mandated mixed-income housing). But what principles can justify intervention into private housing markets for the sake of tenants’ rights to stay put? Such questions take us into some classic questions of political philosophy: Can the state intervene in the economy to bring about more equitable distributions of basic resources? How should states balance the role played by community in securing the good life with the benefits of free markets, which often break up existing communities?

1. Three Approaches to the Wrong of Gentrification

My goal for today is to bring urban geography into conversation with political philosophy in order to argue for two claims: first, that gentrification poses a significant threat to political freedom, and second, that what I call the structural republican approach—not to be confused with the political party of that name—provides the best normative framework to understand why and how gentrification threatens freedom and is thus wrong. According to the structural republican approach, freedom is constituted by a system of legal and socio-economic institutions that allow for the reciprocal co-determination of political life. On this account, gentrification is wrong when and if the influx of new, wealthier residents prevents current inhabitants from having a say in their own lives and meaningful access to the social goods offered by their neighborhoods. The wrong of gentrification is one of unequal political association.

I contrast the structural republican approach with two other approaches. (1) According to the well-being approach, gentrification is wrong because it harms the interests of the poor. (2) According to the communitarian approach, gentrification is wrong because it poses a threat to existing communities and cultural forms of life. I argue that (1) and (2) do not provide adequate grounds to justify coercive state intervention into housing markets.

2. What’s Wrong With Gentrification? Some Preliminary Cases

The examples of gentrification I gave at the outset prime our intuitions that there are in fact problems of justice at stake. In L.A., New York, and San Francisco

²¹² Chester Hartman, “The Right to Stay Put,” *The Gentrification Reader*, ed. Loretta Lees et. al (New York: Routledge, 2010), 531.

²¹³ David Harvey, “The Right to the City,” *New Left Review*, vol. 53 (2008): 23-40.

gentrification has led to eviction, which, given severe housing shortages in these areas, has brought about homelessness, joblessness, poverty, and other social ills clearly at issue for egalitarian theories of justice. For example, the soaring costs of rent in the Bay Area has contributed to a situation in which, shockingly, one in four people in Silicon Valley are at risk of hunger.²¹⁴ But how about cases where the ills have not been quite so severe? Should liberal egalitarians be concerned with those too?

Consider a recent New York Times opinion piece on the gentrification of Harlem. Michael Henry Adams, a long-time resident of Harlem and author of two books on the history of the neighborhood, writes [Q1]:

Gentrification in Harlem might well be likened to the progress of the British Raj, where the most that ‘civilizing’ interlopers could muster was a patronizing interest in token elements of local culture. Thus: Yes to the hip Afro-fusion restaurant, but complaints to 311 over Sundae Sermon dances, barbecues and ball games in parks or church choir rehearsals. These are people who, in saying ‘I don’t see color,’ treat the neighborhood like a blank slate. They have no idea how insulting they are being, denying us our heritage and our stake in Harlem’s future. And, far from government intervention to keep us in our homes, houses of worship and schools, to protect buildings emblematic of black history, we see policies like destructive zoning, with false “trickle down” affordability, changes that incentivize yet more gentrification, sure to transfigure our Harlem forever.²¹⁵

Adams links gentrification to both processes of cultural disrespect and political disempowerment. At stake for him are both the patronizing attitudes of the gentrifiers toward the already-existing cultural forms of the neighborhood—e.g., dances, barbecues, and ball games—and zoning policies that do not give equal weight to interests of current residents. Part of the task of this paper is to urge that it is both philosophically and political useful to keep these arguments separate.

In discourse on the injustice of gentrification, reference to the moral language of rights is ubiquitous. [Q2] For example, Lenina Nadal, communications director for a Brooklyn-based activist network The Right to the City, recommends an “expansion” of the UN Declaration of Human Rights, which “already asserts that everyone has the right to be protected against ‘interference with his... home,’” to the idea that “inhabitants not

²¹⁴ <https://www.theguardian.com/us-news/2017/dec/12/the-silicon-valley-paradox-one-in-four-people-are-at-risk-of-hunger>

²¹⁵ https://www.nytimes.com/2016/05/29/opinion/sunday/the-end-of-black-harlem.html?_r=0

only have a right to their home, a decent, sustainable home...but also to the community they created in their city.”²¹⁶

But caution here is warranted. Against Nadal, one might protest that the culture of neighborhoods stems from decades of cross-cultural mixing and economic diversity. So it is far from a foregone conclusion that anybody has a ‘right’ to a community they have ‘created.’ Presumably, one does not think that Harlem’s Jewish and Italian denizens had community-based claims against the African-American residents who arrived through the Great Migration. It is not hard to understand the frustration of a Latino resident of newly gentrified East Williamsburg in Brooklyn, who concludes, “the neighborhood isn’t supposed to adjust to you... You are supposed to adjust to the neighborhood.”²¹⁷ But what then to make of the fact that East Williamsburg was once an Italian community until Puerto Ricans and other Latin Americans arrived, presumably ‘remaking’ the neighborhood in their own image?

In assessing putative claims of injustice and violations of right we must also remember that most cases of gentrification occur through legal, market-based activity. Imagine someone whose life has been deeply marked by gentrification, a poor woman of color we will name Sally. Due to high rents, Sally now lives two hours each way from work, meaning that she has almost no time to spend with her children. Her old neighborhood had high crime, though it was affordable. But a few ‘adventurous’ urbanites built cafes and bars, even a community garden, in the area. Initially, Sally and the other residents appreciated these establishments, particularly the garden. But over time, these establishments made the neighborhood desirable to more affluent newcomers. This raised rents, which in turn raised property taxes, which in turn meant that Sally’s landlord had very little choice but to raise her rent.

Cases like Sally’s make clear that gentrification often results from routine and entirely legal actions by a number of agents—e.g., renters, home buyers, mortgage lenders, transportation planners, and building and zoning commissioners. Assuming that all of these agents abide by the relevant legal rules such as anti-discrimination laws, assessing the potential injustice of gentrification raises difficult questions about the role of the state in shaping structural outcomes produced by a series of transactions that do not look unjust in and of themselves.

Obviously then it will take conceptual work to understand what, if anything, is so unjust about gentrification that it warrants interfering with landlords ‘rights to property. It is to this task that I now turn.

²¹⁶ <https://www.theatlantic.com/business/archive/2015/09/gentrification-brooklyn-human-rights-violation/402460/>

²¹⁷ <http://www.nydailynews.com/new-york/brooklyn/turf-war-broken-moore-st-east-williamsburg-article-1.1319800>

3. Well-Being

According to the well-being account, gentrification is straightforwardly wrong because it results in a preventable loss of well-being for those displaced. For example, in an influential anti-gentrification article entitled “The Right to Stay Put,” urban planner Chester Hartman writes [Q3]:

Residential stability engenders a host of personal and social benefits. Long-term residence brings safety of person and property...helpful and satisfying social ties to neighbors and local commercial establishments, greater care for public and private space, and lower housing costs...[C]hanges of residence which are foisted on people, which they did not seek out or propose, for which they may lack the social and economic coping resources—these are detrimental to the individual and families involved, and produce social costs as well.²¹⁸

Other studies cite the fact that the displaced must often live further from concentrations of jobs in city centers, which makes it difficult to break the cycle of poverty.²¹⁹ Some court decisions on housing often also involve a similarly well-being based logic. For example, in *Flynn v. City of Cambridge* (1981) the court ruled that the city could grant or withhold eviction permits when the purchasers of a condominium unit want to evict the current occupier. In their decision they stated, “Unless residential rents and eviction of tenants are regulated and controlled, such emergency and the further inflationary pressures resulting therefrom will produce serious threats to the public health, safety and general welfare of the citizens of Cambridge.”²²⁰

The problems with the well-being account of the wrong of gentrification are both empirical and philosophical. Empirically, it is often the case that the early phases of gentrification actually increase diversity, presumably a salient factor in the determination of overall well-being. They may also represent an increase in so-called liberal values, as educated elites challenge the narrative of white flight away from city centers, and the possessive individualist lifestyle of the suburbs. Some studies even show that those who move out of a gentrifying neighborhood to a less-gentrified neighborhood receive better housing for the same rent or equivalent housing for less, so they do not report less

²¹⁸ Chester Hartman, “The Right to Stay Put,” *The Gentrification Reader*, ed. Loretta Lees et. al (New York: Routledge, 2010), 531.

²¹⁹ Hannah Weinstein, “Fighting for a Place Called Home: Litigation Strategies for Challenging Gentrification,” *UCLA Law Review* 794 (2015).

²²⁰ <http://masscases.com/cases/sjc/383/383mass152.html>

satisfaction with their new home. Low-income households that manage to stay often benefit from improvements in the surroundings.²²¹

Philosophically, it is not clear why the well-being of displaced residents grounds a coercive right against gentrifiers, who presumably want to move to a more affordable neighborhood because they feel it will increase *their* well-being. Is the reduction in well-being suffered by the lower middle-class Hispanic resident who now has to move 4 miles further west clearly so much greater than the gain in well-being to the young, white artist couple who face the imminent threat of eviction in their newly yuppified former neighborhood, that the state must step in, whether by imposing rent controls or building mixed-income housing? Not only does it seem nearly impossible to compare well-being in cases like this, our judgments of what constitutes well-being in the first place may be highly moralized (e.g., the good of preserving a minority culture outweighs the hardship of middle-class displacement). Accordingly, enacting policies based on such judgments may come down to state paternalism at odds with liberal neutrality about the good.

4. Community

In a communitarian treatment of these issues Margaret Jane Radin asks why it is acceptable to lower rents at a landlord's expense, a policy that has the effect of favoring current residents over those trying to move to the area. She argues that the well-being approach does not provide an acceptable justification. Interference in private housing markets for the sake of overall well-being represents what Radin calls "nonconsensual governmental alteration of the existing entitlement structure."²²² A libertarian would clearly object to this, but so also would a liberal egalitarian who believes that while the state can determine a basic scheme or pattern of entitlements, it cannot interfere with entitlements that already exist. Indeed, if taken to its extreme, the implication of the well-being account is that property has no self-standing intelligibility, but is merely the effect of a public redistributive scheme.

Radin offers a philosophically more sophisticated version of the "right to stay put," which she calls "the right to stay where you are" (353). She claims that housing is "incompletely commodified." One's home is more closely attached to one's person than the same unit viewed merely as real estate from the landlord's perspective. She writes [Q4]:

Property that is 'personal' ... is bound up with one's personhood, and is distinguishable from property that is held merely instrumentally or for investment and exchange and is therefore purely commercial or

²²¹ This paragraph is indebted to Shaw, "Gentrification: What It is, Why It Is, and What Can Be Done About It." See also, Kelefa Sanneh, "Is Gentrification Really a Problem?" <http://www.newyorker.com/magazine/2016/07/11/is-gentrification-really-a-problem>

²²² Margaret Jane Radin, "Residential Rent Control," Philosophy and Public Affairs, vol. 15, no. 4 (1986): 354-355.

‘fungible.’...[F]ungible property is fully commodified, or represents the ideal of the commodity form, whereas personal property is at least partially noncommodified (362).

On her account, the state has a right to intervene in market-activity so as to prevent gentrification because one’s home expresses the way in which one’s social identity is bound up with the relations of community in which one stands. Anti-gentrification measures are a way of protecting the individual’s right to a group identity.

In my view it is very odd to think that the state has any sort of requirement—even a soft, defeasible one—to preserve a relation to material objects that, for whatever contingent social and psychological reasons, certain people take to be central to their identity. Radin’s view has some surprising and counter-intuitive implications. First, if the reason for protecting people’s homes is that one’s home is where one’s community is, then in fact poorer communities have less of an entitlement to stop gentrification, since poverty often prevents meaningful community from forming in the first place. Those used to continual displacement have less of a personal investment in where they are.

Second, what if the part of physical space I take to constitute my identity-conferring ‘home’ extends out to my ‘neighborhood,’ and my ‘neighborhood’ means ‘the place where people look like me?’ Does the state have any obligation to protect *that*? I am not happy to bite the bullet, as Radin does, and hold that certain forms of exclusionary community should be preserved if its members have themselves been excluded—she cites an Orthodox Jewish community that wishes to ban those who do not observe its religious values or black separatists who wish to exclude whites (380). One need not hold, as Elizabeth Anderson does, that liberal egalitarians stand under “an imperative of integration” to think that the state has no business blocking the efforts of private individuals to integrate, even if those who will be displaced by integration are themselves victims of past discrimination.

In the midst of offering a community-based account of the wrong of gentrification similar to Radin’s, Hartman in the “Right to Stay Put” articulates a different normative justification. He suggests that undoing gentrification promotes “an interest in securing greater equity: more rights and benefits for those in the society who fall at the lower end of the spectrum of resources and power.” It is to such views, which locate the wrong of gentrification less in the effects of gentrification (e.g., loss of well-being, loss of community) and more in the form of unequal power that gentrifying activity represents, to which I now turn.

5. Republicanism: Standard and Structural

I argue that, when it is wrong, the wrong of gentrification can be captured by a prominent philosophical conception of political freedom, according to which an agent is free if she does not have to depend on the arbitrary authority of more powerful agents to pursue her ends. Freedom, on this view, is absence of domination. This is often called the

republican conception of freedom (not to be confused with the political party of that name).²²³

Standard neo-republicanism (SNR)—here I take Philip Pettit as the exemplar—justifies coercive state intervention when it is necessary to prevent unjust asymmetries of power, which it calls domination. SNR defines domination as a “power of interference on an arbitrary basis” (R, 52). For SNR, the paradigm case of domination is the slave whom the master can control at will. Modern cases of the dominated include the wife whose husband can beat her with impunity, and the employee who cannot raise a complaint against his or her employer without threat of termination.

The master, husband, and employer enjoy arbitrary power because they can make choices for the subordinated on bases that have nothing to do with the interests or goals of the subordinated themselves. The decisions of the powerful are not “forced to track what the interests of [the dominated] require according to their own judgments” (R, 55). (2) The master, husband, and employer enjoy an arbitrary power because what matters for domination is not what the dominator actually does but what he *could* do. The essential thought here is that the subordinated are not free if the only reason they can satisfy their desires or interests is that the more powerful are favorably disposed toward them. For example, suppose the slave finds himself owned by a benevolent slave-owner who affords his slaves considerable license. In that case, the slave may be unhindered in most of his choices and so may *appear* to be free. But because this lack of hindrance is due to a factor over which the slave has no control—the continuing goodwill of the master—he is not free, but dominated. Can one simply add “landlord” to the list of dominating agents and thus make sense of the wrong of gentrification?

I think not. In other work, I have argued that while SNR can make sense of *dyadic domination*, it lacks adequate conceptual resources to understand *structural domination*. Dyadic domination is the domination of one individual by another individual. Structural domination is domination by institutions or social forces that are both sustained by and enable instances of dyadic domination. In structural domination, the dominating force is the outcome of uncoordinated actions by a plurality of discrete agents, none of whom may intentionally aim to dominate others. I do not wish to go over those arguments here, except to suggest that because SNR emphasizes cases where one agent wields undue coercive control over another agent, it will understand gentrification through the model of a landlord threatening a tenant with impending homelessness through eviction. Clearly, if a tenant has nowhere to go and is under threat of being priced out, her landlord has a power of interference that does not track her essential interests. According to SNR, rent control or government housing will be necessary in order to constrain such arbitrary power.

But, as we have seen, it is not clear that the case of a landlord threatening a tenant with eviction adequately describes the phenomenon of gentrification. With respect to the scope of gentrifying activity, in Sally’s case the landlord himself is under a host of structural pressures that he cannot control. In some sense then what dominates Sally is a

²²³ See Philip Pettit, *Republicanism: A Theory of Freedom and Government* (Oxford: Oxford University Press, 1999).

pattern of activity and not a person. SNR is forced to locate the threat to freedom that gentrification poses in the terminus of a structural problem—in the way in which large-scale social forces of gentrification are localized in a potential interaction between a landlord and a tenant—rather than in those forces themselves. With respect to the nature of gentrifying activity, gentrification cannot be modeled on eviction—which leaves the tenant with, literally, nowhere to be—in those cases where other housing options are possible.²²⁴

By contrast, according to structural republicanism, gentrification is a problem of freedom not only because it may threaten to subject current residents to the wills of particular landlords, but also because it is a form of asymmetrical disempowerment.

According to structural republicanism, represented by Kant but also in different ways by Rousseau, Fichte, and Hegel, freedom is constituted by a system of legal and socio-economic institutions that allow for the reciprocal co-determination of political life. Here, I follow Niko Kolodny thinking that our concern for domination is ultimately concern for social equality.²²⁵ Individuals are free insofar as they have rights and entitlements, granted and defined by an entire network of norms and institutions, which authorize them to challenge the asymmetrical social power other agents may wield over them. Accordingly, any threat to these norms and institutions, whether stemming from individual actions or structural forces, is a threat to freedom.

This means that a condition of domination is one in which agency is restricted by actions, rules, or structures that the constrained lack opportunity to influence or control, either because: 1.) they cannot meaningfully exit the terms of the relationship or 2.) they lack the socially sanctioned power or standing necessary to co-determine the terms of that relationship. According to this definition, an agent is dominated if she lacks meaningful say in socio-economic and other institutional processes by which her life is constrained. Let me briefly describe each in turn.

Meaningful exit options. A person lacks meaningful exit options when she cannot exit a personal or institutional relationship except on terms that a rational being would not accept. A rational being would not accept terms that necessitate foregoing essential conditions of her agency, including her dignity, self-respect, and standing as a social equal. One potential rule of thumb in ascertaining the rational acceptability of terms is whether dominating individuals (or individuals acting in roles empowered by dominating institutions) would accept those terms should they find themselves in the position of the dominated.²²⁶

²²⁴ See here Arthur Ripstein, *Force and Freedom* (Cambridge: Harvard University Press, 2009), 232-267; Jeremy Waldron, “Homelessness and the Issue of Freedom,” *UCLA Law Review* 39 (1991): 295-324; and Christopher Essert, “Property and Homelessness,” *Philosophy and Public Affairs*, vol. 44, no. 4 (2017): 266-295.

²²⁵ See Kolodny, “Being Under the Power of Others” (unpublished).

²²⁶ For a similar emphasis on the lack of reciprocal acceptability in dominating interactions, see Rainer Forst, “A Kantian Conception of Justice as Nondomination,” in *Republican Democracy: Law, Liberty and Politics*

Co-determining powers. In many cases the very idea of exit options is incoherent, since opting out would necessitate forgoing the essentials of a decent life in the modern world (e.g., employment, housing, health care, etc.) In those cases, a person is dominated if she lacks the socially sanctioned power or standing necessary to co-determine the terms of the relationship.

In requiring meaningful exit options and co-determining powers, structural republicanism does not require that in order to be politically free I must have the opportunity to meaningfully determine the role of *every* market-based interaction in which I engage. It does not imply that I am unfree whenever market pressures push me to seek a change of residence. But it does demand that I have opportunities to affect the basic terms or patterns of such interactions.

So in assessing whether gentrification dominates an agent, one needs to ask: does she have a role in the processes by which rental market conditions are set? Can she petition the government to build public housing or to control rents? Will her voice be heard, or will the interests of various lobbies drown her out? ? In cases where actual co-determination proves impossible, giving the subordinated equal voice might have to be approximated through various legal measures. So, in determining whether the agent is dominated, we might also ask, is the landlord legally constrained in various ways? For example, are there limits on how much he can increase the rent and is he mandated to keep a certain number of units under rent control, or is rent entirely at his discretion, even if sudden hikes threaten to leave many tenants destitute, if not homeless? Can the landlord simply force renovations on tenants, thereby raising the value of the apartment and thereby possibly exempting the unit from rent control—as is current practice in cities like New York?²²⁷

On the structural republican approach, rent controls and restrictions on landlords will be justified as the best way of representing the interests of the neighborhood's current residents in the collective co-determination of public life.

Conclusion

I have argued that gentrification is wrong when and if it threatens the rights of both former residents and newcomers to collectively co-determine the terms of their lives. On my account, issues of who was there first, or rights to something called community, or considerations of overall well-being, play no role in the account of the wrong of gentrification.

(Edinburgh: Edinburgh University Press, 2013): 154-168, and Vrousalis, "Exploitation, Vulnerability, and Social Domination": 140-141.

²²⁷ See Michael Greenberg, "Tenants Under Siege: Inside New York City's Housing Crises," New York Review of Books, August 17, 2017.

PITTING PEOPLE AGAINST EACH OTHER

Waheed Hussain
University of Toronto
waheed.hussain@utoronto.ca

Waheed Hussain is an assistant professor of philosophy at the University of Toronto. His interests lie mainly at the intersection of political philosophy, economics and business. He was a Visiting Faculty Fellow at the University Center for Human Values at Princeton in 2013-14. He has published papers in academic journals such as Philosophy & Public Affairs, Economics and Philosophy, and the Journal of Business Ethics, and he is currently working on a book about markets and freedom.

Many philosophers believe that the central requirements of political morality may be understood in terms of two fundamental ideals: liberty, on the one hand, and equality or fairness, on the other. If other ideals are important they are important mainly because they set out effective strategies for maintaining institutions that satisfy the demands of liberty and equality. My goal in this article is to show that there is an aspect of political morality that transcends liberty and equality. The aspect that I have in mind corresponds to the third element in the great triad of liberal democratic thought: Liberty, Equality, Fraternity.

The main idea might be put like this. A society's basic institutions can mistreat people in various ways: for example, institutions can fail to provide people with certain fundamental rights and freedoms, or they can unfairly advantage some people over others. But in addition, institutions can mistreat people by creating social circumstances in which the only way for one person to secure an important good is by formulating and successfully carrying out a plan that will effectively interfere with another person's formulating and successfully carrying out a plan to secure an important good. That is, institutions can create circumstances in which people have to struggle against each other, or as I will put it, institutions can "pit people against each other."

Pitting people against each other is closely related to competition. Many social institutions tend to be competitive, including labor markets, college admissions processes and democratic elections. A presidential election, for example, typically creates a situation in which for any one candidate to secure the office, she has to formulate and carry out a plan that will effectively block all of the other candidates from securing the office. Competitive institutions have many things going for them, as they can encourage people to develop their talents, place qualified candidates in important positions, and allocate productive assets to their most socially beneficial uses. Competition is often a good thing and no plausible account of political morality would rule out competition altogether.

It is equally clear, however, that there are moral limits to competition. Suppose that a liberal democracy is considering a proposal to weaken its public health care system.²²⁸ Instead of providing health care to individuals free of charge and bearing the costs collectively, it would be up to each citizen to secure a market income and buy private insurance for herself and her dependents. If we assume that something short of full employment is the normal situation, moving health care into the sphere of market distribution would put citizens on a very different footing with one another. Since there are more job seekers than jobs, each citizen would find that in pursuing job opportunities, partly to secure health care for herself and her family, she must carry out a plan that would effectively block some other citizen from getting a job and securing health care for herself and her family. Citizens would effectively pose an active threat to one another's basic health care coverage.

The mutual threat in the health care case may seem relatively benign when we view it from the detached perspective of an economist or a bureaucratic planner. It may even be attractive insofar as it gives people an incentive to surpass one another in terms of skill and productivity. But the mutual threat is much less benign when we consider it from the perspective of the human beings who have to live together under the new arrangement. In a high stakes labor market, each citizen looking for a job knows that her activities will do very real damage to the citizens whom she beats out, but she can't act on her concern for others if she's going to land a job and prevent the damage from being done to herself and her loved ones. She must therefore act with a kind of disregard for her fellow citizens. They in turn must act with a kind of disregard for her. And mutual disregard is, in this way, built into the nature of intensely competitive institutions.²²⁹ With the general decline in funding for public health care, education and other services in recent years, as well as the reduction in barriers to international labor mobility, many citizens in liberal democratic societies today find themselves caught up in an increasingly intense, worldwide struggle to secure employment and the various goods attached.

In what follows, I will develop a theory that can explain both why competitive institutions are permissible in a liberal democracy and why there are moral limits. Many philosophers recognize that citizens in a liberal democracy stand in a special relationship with one another (Walzer 1983; Dworkin 1986; Scanlon 2003b; Rawls 2005b; Cohen 2009; see also Rousseau 1979 and 1997a). According to what I call the *fragmentation account*, this relationship puts certain constraints on the competitive character of social institutions. In much the same way that a properly ordered poker game among friends must stay within the bounds of a "friendly competition," a properly ordered scheme of institutions in a liberal democracy must keep the competitive character of the

²²⁸ As I write this article, a third attempt to repeal the Affordable Care Act (ACA) has failed in the US Senate. The Graham-Cassidy bill would have taken over \$800 billion out of the Medicare program and dismantled the ACA's regulations, taxes and public subsidies at the national level. These measures would have made access to health care depend much more significantly on how an individual (or members of her family) fares in labor market competition. More repeal efforts are on the way.

²²⁹ There may be other moral defects in the high stakes labor market (e.g. it does not satisfy the basic right to health care). But pitting people against each other is clearly one among its defects, and we cannot formulate a complete theory of institutional morality without paying attention to this defect.

arrangement within the bounds appropriate to a community whose members should care about the common good. My main point is not that citizens are required to care about one another, but rather that social institutions must not be excessively hostile to their relating to each other in this way. A liberal democracy may adopt competitive institutions, but it must keep these arrangements within the proper bounds.²³⁰

Ultimately, the goal of the fragmentation account is not to motivate some form of socialism, but rather: (1) to articulate the moral appeal of social democratic arrangements that incorporate markets and liberal freedom, but moderate and contain the sphere of competition in social life, and (2) to show that the moral appeal of these arrangements is not fundamentally tied to equality, whether luck or relational.

1. Rivalry Defining Arrangements

Philosophers have not squarely addressed the issues that I want to discuss, so I begin with some basic ideas.²³¹

A social institution is a collective activity defined by a set of rules. Members of a community treat the rules as authoritative standards for how to think and act in certain circumstances, and they appeal to the rules in assessing one another's thought and conduct (Hart 1991; Rawls 1999; Scanlon 2003a). In the cases that matter for my purposes, the rules of an institution will define various "moves" that members can use to affect their own institutional situation and that of other members, and members are required to formulate their various projects within the long-run possibilities available in the arrangement. For example, consider a course-scheduling scheme at a university. The rules of the scheme define various moves for members—e.g. professors can propose classes, creating new options for students; and students can add or drop classes, shaping the institutional responsibilities of professors. Members of the university community form their various plans for what courses to teach and what classes to take within the scheme of long run possibilities available given the interaction defined by the scheduling process.

²³⁰ Even Rousseau (1960, 126-7; 1979, 352-4, 141-2; 1997b, 191) gives competitions a place in the ideal republic.

²³¹ There is a well-developed literature on how people should act within adversarial institutions (e.g. Applbaum 1999). But this literature does not directly address the morality of establishing these institutions in the first place. McMahon 1981 and Gomberg 1995 address issues more closely connected to the latter question.

Social institutions can affect the availability of goods in a community and they can structure access to these goods. Among the ways that institutions affect availability is through incentives: for example, institutional rules at a university can increase the number of spots available in Calculus 101 by offering professors incentives to teach more sections of the class. Among the ways that institutions structure access is by attaching claims on goods to an institutional status and then defining a process for securing the status. In the case of Calculus 101, for example, the rules of the university may attach institutional claims on attending the class to the status of being “registered” for it, and the institution may then define a process for students to register.

Some social institutions have the property of being *rivalry defining*. Call an institution that has this property a *Rivalry Defining Arrangement* (RDA). The central idea is that the rules of an RDA generate and structure access to goods in such a way that the members of a community will find themselves in a certain kind of strategic situation. More formally:

Rivalry Defining Arrangement (RDA): A social institution is an RDA when: (1) it generates and structures access to goods in such a way that, other things being equal, a community of rational individuals²³² treating the rules as an authoritative framework for thought and conduct will find that there are at least two members of the community, *A* and *B*, such that *A*'s formulating and successfully carrying out a plan to realize *A*'s aspirations within the framework of the rules would interfere with *B*'s formulating and successfully carrying out a plan to realize *B*'s aspirations within the framework of the rules; and (2) the rules do not direct anyone to take any special measures to conceal the strategic situation from those involved.²³³

A tennis match is typically an RDA. The rules of the enterprise attach claims to recognitional goods (e.g. trophies, applause) to the institutional status of “winning.” The rules then define a process for securing this status. When everyone treats the rules as authoritative, including audience members and officials, they create an environment in which, other things being equal, rational individuals will find themselves in a certain strategic situation. Each player will aspire to win. But according to the rules there can only be one winner, determined by who wins each point and takes the right number of sets. So each player finds that if she formulates and successfully carries out a plan to realize her aspiration, this would effectively block the other player from doing the same thing. Or, to put it another way, each player finds that the other is a threat: if her

²³² For my purposes, a rational individual is someone who forms and revises her intermediate goals appropriately in light of her final ends, and forms and revises her final ends appropriately in light of the availability of various objective human goods.

²³³ An arrangement that pits people against each other but seeks to conceal this fact may also be morally defective, on my view. But the analysis of this defect is more complicated and requires a more extensive discussion of the way that full publicity defines the proper circumstances in which to assess social institutions.

opponent formulates and successfully carries out a plan to realize her aspiration this will block the player herself from doing the same thing.²³⁴

The property of being rivalry defining is scalar. An institution is more or less intensely rivalry-defining depending on the strategic situation that would result, other things being equal, in a community of rational agents who adhere to the scheme. Under a more intensely rivalry defining arrangement, two representative individuals *i* and *j* would find that *i*'s formulating and successfully carrying out a plan to realize her aspirations would (1) interfere with more of *j*'s pursuits, (2) interfere with pursuits that are more important to *j*, or (3) come closer to blocking *j* from achieving something she cares about rather than just making it harder for *j* to achieve it. For instance, a tennis match would be more intensely rivalry defining if the rules say that the winner lives and the loser dies. The reason is that the players would threaten each other with respect to something that is worth more to each of them, namely their lives.

Some readers might assume that any institution that distributes goods that are scarce must be an RDA. This is not true. The rivalry defining character of an institution is never simply a function of background facts about material scarcity. Suppose there are 100 people in a community who need a kidney transplant and only 10 kidneys available. There is nothing anyone can do either to reduce the need or to increase the supply. In this case, a social institution could address the situation by defining a lottery (or some similar scheme²³⁵). The lottery arrangement would attach institutional claims on kidneys to a certain status and then assign the status to 10 people according to a set of probabilities. But it would not create any institutional avenue for one individual to formulate and carry out a plan that would reduce anyone else's chances of securing the relevant status. Insofar as the arrangement does not put people in a situation where they could affect one another's institutional claims, it does not translate brute facts about scarcity into a social situation in which people pose strategic threats to one another.

Framing the issue in the way that I do presupposes a certain nonconsequentialist approach to institutional morality. The rules of an institution define a context for social interaction, and this context may be morally significant for reasons that are not simply tied to the outcomes that it will generate. Imagine that two communities, X and Y, each face the kidney shortage that I described above. X structures access to kidneys through a lottery scheme. Y structures access through a tennis tournament, where the top 10 finishers get a transplant. The same underlying conflict of interests exists in each society, but the institutional arrangements in each case are different. The lottery scheme does not allow one member's actions to affect another member's chances of securing a kidney, so this arrangement does not pit people against each other in my sense. By contrast, the tennis tournament does. Even if the two arrangements generate identical distributions of kidneys, the difference in the way that they situate people relative to one another may be

²³⁴ Institutional competitions are one type of RDA. I discuss another type in section 8.

²³⁵ I discuss some other possibilities in section 8.

morally significant because situating people this way violates a nonconsequentialist ideal for social institutions.

Three further points about RDAs will be important for my discussion. First, RDAs do not ordinarily require members of a community to form an intention to harm or defeat one another. In the case of a tennis match, each player only has to form an intention to hit every ball into the opposing court: if she gets every ball back, she wins. The deprivation suffered by the other player need only be a foreseeable side effect of each player's pursuit of the valuable status defined by the institution.

Second, moderating the rivalry defining character of an institution does not require that the institution foster any form of cooperative interaction. An institution would be more moderate in its rivalry defining character if it were structured in such a way that, other things being equal, one person's achieving her aims (1) would contribute positively to other people's achieving their aims or (2) would simply not interfere with other people's achieving their aims.

Third—and most importantly—the institutions that are my main concern in this article are *engulfing arrangements*. These institutions have profound effects on people's lives and people are not in a position to opt out of them on a voluntary basis. A game of tennis is usually a relatively contained affair: it has a limited impact on people's lives and each player can decide to stop playing if she likes. But large-scale social institutions are not like this. Consider the rules of ownership, exchange, inheritance and public provision²³⁶ in a liberal democracy. Together, these rules shape the distribution of institutional claims that each citizen confronts when she is born and they shape what each citizen must do to secure institutional claims on food, shelter and medicine. The rules apply to everyone as a matter of law and no individual can decide to “stop playing” on a voluntary basis. The difference between games and large-scale institutions is extremely important, as certain properties that may be acceptable in the context of a game are more problematic in the context of an engulfing arrangement.

2. Morally defective RDAs

²³⁶ By “public provision” I mean publicly funded schemes that provide more or less universal access to a good in a community. Some examples in contemporary liberal democracies include arrangements that provide universal access to: running water, roads and sidewalks, primary and secondary education, health care (e.g. Medicare, NHS), police protection, courts and the judicial system, sports and recreation facilities, libraries, museums, parks and public spaces, retirement income, information about current events (e.g. BBC, CBC), statistical data (e.g. census), scientific knowledge, sewers and sanitation services, and so on.

The institutional order in a modern liberal democracy is an engulfing arrangement. Citizens are born into an order with certain rules of ownership, exchange, inheritance and public provision. This order has profound effects on people's lives and individuals are generally not in a position to opt out on a voluntary basis. The question I want to ask is, does the fact that an engulfing arrangement is an RDA count as a moral defect in the regime? In this section, I want to show that it sometimes does.

To examine the issue, I will consider an example where an institution has profound effects on people's lives and they are not in a position to opt out on a voluntary basis. Over the course of this article, I will show that the example is a good model for thinking about the institutional order in a modern liberal democracy.

Here is the example.

Physics Department. A physics department has two tenure-track assistant professors, *A* and *B*. Each one works in a separate subfield and has a position with a separate "line." According to the rules, *A* will get tenure if she makes a significant contribution to the field, regardless of what happens to *B*. And *B* will get tenure if she makes a significant contribution to the field, regardless of what happens to *A*. The department gives no one any guarantee that the rules won't change, just as many other department policies might change. Given their talents and the state of the field, *A* and *B* each has a 50% chance of getting tenure.

One day, the department institutes a new arrangement. The new arrangement collapses the two lines into one, so either *A* or *B* will get tenure, depending on who makes the more important contribution. If *A* gets tenure, *B* will not. If *B* gets tenure, *A* will not. Given their talents and the state of the field, *A* and *B* each has a 50% chance of getting tenure.

Call the first tenure scheme "S1" and the second "S2." For the purposes of my discussion, I stipulate that both schemes are equally feasible, so with respect to S1 especially, the department has the resources to tenure both candidates if they both satisfy the relevant tenure standard.

I believe that there is a distinctive moral defect in S2 as compared to S1. Let's assume that both arrangements are structured so that, other things being equal, rational people in the position of junior faculty members will aspire to become tenured professors. S1 gives *A* and *B* "clear pathways" to tenure. *A* can formulate and successfully carry out a plan for realizing her aspiration without interfering with *B*'s prospects. And *B* can formulate and successfully carry out a plan for realizing her aspiration without interfering with *A*'s prospects. This feature of the arrangement is important from the standpoint of social interaction. When the relevant facts come to *A*'s attention, *A* can share in *B*'s aspirations, hoping that *B* succeeds in her research program and helping *B* to achieve her

aims because *B*'s success in this regard does no damage to *A*'s prospects. Similarly, when the relevant facts come to *B*'s attention, *B* can share in *A*'s aspirations, hoping that *A* succeeds in her research program and helping *A* to achieve her aims because *A*'s success does no damage to *B*'s prospects. The agent relative reasons that each faculty member has for caring about her own projects do not give her any special reason not to share in the aspirations of her colleague.

By contrast, S2 does not give *A* and *B* “clear pathways” to tenure. For *A* to realize her aspiration she has to formulate and successfully carry out a plan that will effectively block *B* from doing the same thing. And for *B* to realize her aspiration, she has to formulate and successfully carry out a plan that will effectively block *A* from doing the same thing. This creates a very different environment. Suppose that *A* starts producing more than *B*. With every paper that *A* publishes, she pushes *B* one step closer to failure. But if *A* wants to realize her aspiration, she has to go on publishing, one paper after another, steadily running *B*'s dream into the ground. It follows that *A* has a powerful reason under S2 to act with a certain disregard for *B*'s aspirations. *A* has a powerful reason not to be moved by the setbacks that she imposes on *B* (since this might make *A* tentative and reluctant) and a powerful reason not to help *B* in her research projects.

Something similar plays out from the loser's perspective. Suppose that *A* starts producing more than *B*. If *B* wants to realize her aspiration, she has to pull herself together and get back in the game. But any step that *B* takes towards realizing her aspiration requires her to do something that pushes *A* one step closer to failure. Given that *B* must undermine *A*'s aspirations as a necessary side-effect of her own pursuits, she must act with a certain disregard for *A*'s projects. So *B* also has a powerful reason not to be moved by the setbacks she imposes on *A* and a powerful reason not to help *A* in her projects.

As I noted earlier, it may be morally acceptable for institutions to be rivalry defining in certain kinds of situations—e.g. the circumstances of the typical tennis game. But the department's tenure system is an engulfing arrangement: it has profound effects on people's lives and people cannot opt out on a fully voluntary basis. The tenure scheme is not “just a game,” and it is partly in virtue of this fact that the rivalry defining character of S2 constitutes a moral defect in the regime.

An analogy might help to crystallize the idea. Suppose that a Roman emperor throws two citizens into a gladiatorial arena. There is only one exit. The emperor does not tell anyone what to do, but announces a rule: “two men enter; one man leaves.” As the first citizen starts towards the door, he knows that he is effectively damaging his fellow citizen's prospects for escape, but what can he do? When the second citizen sees the first heading for the door, instead of welcoming the possibility of his compatriot escaping, he starts running for the exit himself. As they both race to the door, they are each responding to the very powerful reasons that the institution creates for them to act with a kind of disregard for the other person and his projects. As some would say, the arrangement sets

the stage for “man’s alienation from his fellow man.”²³⁷ Now I take it that the emperor has serious moral reasons not to treat people this way and that the reasons go beyond the pain and suffering he will indirectly inflict on those involved. Part of the reason has to do with the fact that the emperor would be putting human beings, who are capable of acting with evident concern for one another, in an environment where they have powerful reasons to act with intense mutual disregard. Under S2, the physics department would be throwing its junior faculty members into an academic cage match, and it has a serious moral reason not to do this.

Let me add that the moral defect in S2, as I understand it, lies in the structure of the arrangement itself, rather than the motivations of those who adopt it. If the members of the department were to come together to discuss changing their tenure scheme from S1 to S2, any faculty member could cite the fact that S2 has a certain structure to show that the arrangement pits junior faculty members against each other and is therefore defective. Whether the members of the department adopt S2 with the intention of pitting junior faculty members against each other may affect our moral appraisal of these individuals, but it does not bear on the defectiveness of the institution itself (see Scanlon 2008).

3. Good feelings cannot be the whole story

There is a distinctive moral defect in S2 as compared to S1. I will set out the correct explanation of this defect in section 6 below. But in order to motivate my view, I want to show how several ideas in contemporary political philosophy cannot provide a satisfactory explanation on their own.

Start with good feelings and dispositions. Suppose that it is intrinsically good for human beings to have friendly feelings and supportive dispositions towards one another. Given this assumption, some might argue that the distinctive problem with S2 is that the feelings and dispositions likely to develop under S2 are worse than those likely to develop under S1.

The intrinsic goodness of certain feelings and dispositions may explain why some RDAs are morally defective, but it cannot explain the defect in S2 because the issue here is not simply a matter of the actual feelings and dispositions that would develop under the arrangement. Here is an illustration.

²³⁷ Though Marx (1964) uses the phrase, he has something different in mind. Marx wants to draw attention to the authoritarian character of market society, where members create market arrangements but then become subservient to the impersonal drive for capital accumulation built into the market process. Subservience to the impersonal drive leads to a more specific and personal form of subservience to the particular individuals (i.e. capitalists and their managers) in whose activities the accumulation process is partly realized.

Physics Department (b). The department in the original example is considering a move from S1 to S2. This time, *A* and *B* happen to be good friends when they are hired. Their friendship is so strong that, as a matter of fact, *A* and *B* will not slide into indifference or hostility under S2, but will remain on friendly terms.

Given the circumstances in Physics Department (b), there is no difference between S1 and S2 in terms of the feelings and dispositions that will develop under each arrangement. Nonetheless, there is still a distinctive moral defect in S2. The point might be clearer if we imagine again that the members of the department meet to discuss the move from S1 to S2. Some members might argue that since there will be no net change in people's feelings and dispositions under S2, there is nothing morally objectionable about the arrangement. But other members would be correct to reject this line of reasoning. Even if the friendship between *A* and *B* would prevent certain feelings and dispositions from emerging under S2, the arrangement would still mistreat junior faculty members. S2 creates a situation in which *A* and *B* have very powerful *reasons* to act with disregard for the realization of one another's projects. The fact that S2 embodies this form of treatment is itself a failing in the institution, even if creating these reasons does not actually generate the corresponding attitudes and behaviors. If I were in the position of *A* or *B* in the faculty meeting, I would be right in thinking that the new arrangement mistreats us and the fact that she and I happen to be good friends does not give the department a license to pit us against each other.

4. Fairness and equality cannot be the whole story

Another possible explanation of the moral defect in S2 appeals to fairness: some might argue that S2 is unfair in a way that S1 is not. This view is not plausible, however, because there is no obvious difference between the two tenure schemes with respect to fairness. Neither arrangement tilts in favor of one of the candidates rather than the other, and in both cases, the candidates are assessed in terms of relevant criteria, such as the quality of their work and their service to the department. The two arrangements are evenhanded in their treatment of the candidates.

What about equality? The two arrangements provide junior faculty with similar tenure prospects and similar resources, so there is no difference between the two regimes in terms of either equality of opportunity or equality of resources. In terms of equality of outcomes, however, there is an important difference. S1 allows for the possibility of an equal outcome—i.e. *A* and *B* could both get tenure or both get denied tenure. But S2 will *necessarily* lead to an unequal outcome, as one will get tenure and the other will not.

Some might argue that this inherent tendency towards inequality can explain the distinctive moral defect in S2.

We can reformulate the example, however, to show that equality of outcomes cannot be the whole story.

Physics Department (c). The department in the original example has no tenure scheme and is choosing between two possibilities. Under S3, the department secures approval for two “lines” and each candidate will get tenure if she satisfies the tenure standard, regardless of what happens to the other candidate. The standard is based on research and teaching, and given the facts, *A* and *B* each have a 50% chance of getting tenure.

Under S4, there is a competition: if one candidate surpasses the other in terms of both research and teaching, that candidate alone will get tenure. If no candidate does better along both dimensions, the department will seek approval from the dean for two “lines” in order to tenure both candidates, but there is a 50% chance that the dean will say yes and a 50% chance that she will reject both candidates. The candidates are evenly matched, so they each have a 50% chance of doing better in terms of research and a 50% chance of doing better in terms of teaching.

The example is a little more complicated, but much as in the original example, there is a distinctive moral failing in S4 as compared to S3. Under S3, *A* and *B* have “clear pathways” to tenure: they can each formulate and carry out a plan for getting tenure without interfering with the other candidate’s prospects. Under S4, the candidates do not have “clear pathways” to tenure. According to the rules, each candidate has two basic routes to tenure: (1) surpass the other candidate in terms of both research and teaching, or (2) surpass the other candidate along one dimension and hope for a favorable decision from the dean. If *A* publishes a paper or gets a good teaching evaluation, this effectively constitutes a setback for *B* in terms of both routes: *B* has to publish one more paper or get one more good teaching evaluation just to stay even. Similarly, if *B* publishes a paper or gets a good teaching evaluation, this effectively constitutes a setback for *A* in terms of both routes. As such, S4 pits *A* and *B* against each other in a way that S3 does not. Notice, however, that there is no difference between the two schemes in terms of the probability of an equal outcome: in both cases there is a 50% chance, i.e. a 25% chance that both will get tenure and a 25% chance that both will get denied tenure.²³⁸

No doubt, there are more elaborate ways of trying to explain the moral defect in certain engulfing RDAs by appeal to fairness and equality, but these accounts are also

²³⁸ Under S3, there is a 50% chance of an equal outcome (i.e. a 25% chance that both candidates will get tenure and a 25% chance that both will get denied tenure). Under S4, there is a 50% that no one will win outright at stage 1, at which point there is a 50% chance that the dean will tenure both candidates and a 50% chance she will deny them both. So at the outset, there is a 50% chance of an equal outcome (i.e. a 25% chance that both candidates will get tenure and a 25% chance that both will get denied tenure).

likely to be inadequate. The reason is that fairness and equality are values that apply to an institution most naturally insofar as it distributes benefits (or burdens) to some people rather than others. But the distinctive moral defect in S2 is not fundamentally about where the benefits and burdens end up. The defect is fundamentally about *fighting*: the second tenure scheme fashions an environment in which people must foreseeably crush each other's dreams to get ahead. The arrangement creates a certain environment for social interaction, and we need an account of the moral failing that is more closely tied to this concern.

5. Face to face interaction is not essential

Before developing a better account of the moral problem with engulfing RDAs, I want to address a different kind of objection. Some readers might accept that engulfing RDAs are morally problematic in the context of a small department where people meet each other face to face. But they might think that this does not have implications for the engulfing institutions of a liberal democracy because citizens in a modern society do not meet the vast majority of their fellow citizens face to face.

Although certain moral defects in an institution may presuppose face to face interaction, this is not the case when it comes to the moral defect at issue in this article. Here is an illustration.

Physics Department (d). The original physics department is once again considering the choice between S1 and S2. This time, however, the department occupies two floors in a large academic building. People on different floors never meet each other face to face. Junior professor *A* is on the higher floor and junior professor *B* is on the lower one. *A*'s office happens to be directly above *B*'s, and because of the ventilation system, *A* can hear joyous whoops when things go well for *B* and quiet sobbing when things go badly for her. The same holds true for *B* with respect to *A*.

I take it that there is a distinctive moral defect in S2, even in the circumstances set out in Physics Department (d), and that it does not make a difference whether the junior faculty members meet face to face. Think of it this way. Though *A* and *B* never meet, they know perfectly well the structure of the second tenure scheme and how their pathways to tenure would affect the other person, abstractly conceived. Moreover, when *A* hears joyous whoops from downstairs or quiet sobbing, she understands in a clear and

present way the impact that her pursuits have on the other person. The same holds true for *B* with respect to *A*. So although they never meet face to face and have no personal acquaintance, the institutional arrangement still gives them reasons to act with a certain disregard for the other person and her projects. Under these circumstances, when the choice between *S1* and *S2* comes up in a faculty meeting (held concurrently on the two floors), *A* and *B* would each be correct in pointing out that it is morally objectionable to pit them against each other. What is essential to the objection is the way that the institution situates people relative to one another; whether there is face to face interaction is beside the point.

In order to make the issues more concrete, let me say something about one way that we encounter the institutional structure I have been discussing. In a modern liberal democracy, social life consists, among other things, of a vast array of competitions for attractive positions, including competitions for jobs, university admissions, and places in good secondary and primary schools. All of these competitions exist, in part, because institutional rules create the conditions for competition. For example, labor market competition exists in part because the law gives employers the legal right to hire and fire employees for most any non-discriminatory reason. The rules of taxation and public provision define the stakes in the competition by determining how much a citizen's access to basic resources will depend on how she fares in the labor market. And the legal rules of employment, immigration, social security, pensions, corporate governance, and so on, further define the competitive environment.

The rivalry defining character of the institutional rules creates reasons for citizens to act with a certain disregard for one another. This disregard can take the form of attitudes directed at people identified as particular individuals, but it can also take the form of attitudes directed at people identified in terms of abstract categories. For example, the institutional order in most liberal democracies today puts native born workers and recent immigrants in a high-stakes struggle for jobs, education, housing and social standing. With regard to any particular job or position, people may not know the other individuals they are competing with. But at a social level, everyone understands certain basic facts: native born workers correctly perceive recent immigrants as a threat to their resources and social standing, and recent immigrants correctly perceive native born workers as a threat to their resources and social standing. Each citizen knows that she must act with evident disregard for the interests of others or else get pushed into the dirt herself. As the struggle unfolds, fear, anger, and resentment may course through the body politic, but these attitudes will generally take the form of attitudes directed towards people conceived in terms of abstract categories—e.g. *native born*, *immigrants*, *illegals*, etc.—rather than as particular individuals.²³⁹

Though this is not often stressed, society's basic institutions play an important role in generating the pattern of attitudes I just described. The rivalry defining character of basic institutions in many liberal democracies puts people in a situation where they

²³⁹ Various factors (e.g. history, proximity, social ties, etc.) can lead people to conceive of the conflict in terms of particular abstract categories, such as race or ethnicity. People may conceive of the conflict in these ways even if social institutions actually pit people against each other as atomized individuals.

pose active threats to one another. On my view, when citizens come together in the political forum, they can each correctly point to the intensely rivalry defining character of their basic institutions as a serious moral defect in the arrangement, and it makes no difference that each individual citizen does not meet all of her fellow citizens face to face.

6. The fragmentation account

Engulfing RDAs are sometimes morally defective in virtue of their rivalry defining character. Face to face interaction and personal acquaintance are not essential to the problem, so there is no special reason why the institutions of a modern liberal democracy cannot suffer from the same defect. I considered good feelings, fairness, and equality, and none of these ideas offers a full explanation. How then should we understand the problem? According to the fragmentation account, engulfing institutions are morally defective in virtue of their rivalry defining character when they are antithetical to particular forms of solidaristic integration among individuals.

Solidaristic integration is a way for two or more people to relate to each other. Relating to each other in this way involves each person giving a subset of the projects of the others a status in her practical reasoning that is analogous to the status that she gives to her own projects in her practical reasoning. We might think of this as a particular way for people to care about each other.²⁴⁰ For example, suppose that I live in a neighborhood with *A* and *B*, and that we are solidaristically integrated with respect to each person's project of maintaining a home for her family. Suppose that a brushfire threatens to destroy *A*'s house. Insofar as we are integrated, when the relevant facts come to my attention, I will react to the situation as if the brushfire, in threatening *A*'s efforts to maintain a home for her family, were also threatening my efforts to maintain a home for my family. Given our solidaristic integration, *B* will react in a similar way. So, when *B* and I see the brushfire, among other things, we will worry about it damaging *A*'s house, pitch in to help *A* to save her house, and we will be relieved when the brushfire is put out.

²⁴⁰ Classic accounts of solidaristic integration include: Plato, *Republic*, 462b-d; Aristotle, *Nicomachean Ethics*, book 9; Rousseau 1979 and 1997a; and Kant's (1996) duties of mutual love and mutual respect. All of these accounts differ in some ways from mine. A non-solidaristic form of integration involves two or more people who see themselves as playing complementary parts in an encompassing enterprise that they each value as a final end—see Rawls's discussion of "social union" (1999, 456-64). This latter form of integration may also be valuable, but it is not central to understanding why certain RDAs are morally defective (see [name withheld] 2017).

Many social relationships require a form of solidaristic integration from those who stand in these relationships.²⁴¹ The most obvious case is friendship: if you and I are friends, our relationship requires that I give some subset of your projects a status in my practical reasoning that is analogous to the status of my own projects in my practical reasoning. I should hope for your success in these projects, be anxious about your failure, and try to help you to realize these aims. The same applies to you with respect to me.

Another case involves academic colleagues. Members of an academic department stand in a certain relationship with one another, and this relationship requires a form of solidaristic integration, i.e. professional collegiality, which is centered on professional projects. Among other things, collegiality requires that when a colleague finds out about a fellow colleague's difficulties in a work project, she should share in her colleague's frustrations and be open to helping her. And when she finds out that her colleague has solved the problem, she should feel relief and satisfaction.

For my purposes, the most important relationship that requires a form of solidaristic integration is the political relationship. As I noted earlier (p.4), many philosophers hold that the members of a political community stand in a special relationship with one another. On my view, the political relationship is a kind of partnership, one that requires a form of solidaristic integration among citizens. We can think of the relevant form of integration in terms of a conception of the *common good* ([name withheld] 2017; 2018). On my liberal view, the common good consists in a set of resources that serve or could serve most any citizen's interest in being able to form, revise and pursue her own conception of the good. These resources include, among other things, civil and political liberties, income, health care, housing, and a public basis for self-respect. What the political relationship requires is for citizens to be solidaristically integrated with respect to one another's efforts to secure the elements of the common good for themselves. For example, if the members of a minority group are struggling to secure their basic liberties against a threat, the political relationship requires that I, as a fellow citizen, should regard the realization of this project as something to hope for and something to help members of the minority group to achieve. And it requires that I should regard the possibility of their failure as something to be anxious about and something to help them to avoid.

Solidaristic integration is a pattern of activity, a distinctive way for people interact with one another. Social institutions can obstruct this form of interaction. We can think of obstruction here in terms of reasons for *distancing*. A person *A* "distances" herself from a person *B* when *A* does not give *B*'s projects the status in her reasoning that is required by their relationship. Distancing can take the form of indifference, where *A* treats *B*'s success or failure in a certain project as not warranting any particular attitudes or actions. But distancing can go beyond indifference into enmity, where *A* treats *B*'s success or failure as warranting attitudes and actions that are the opposite of what would be required

²⁴¹ A normative relationship is a set of reciprocal requirements that apply to two or more people when they satisfy a certain set of conditions with respect to one another. Individuals "stand in" a certain normative relationship when they satisfy the relevant conditions. See Dworkin 1986, 195-216; and Scanlon 2008, 131-8.

by the relevant form of integration (e.g. A hopes that her colleague will fail in a professional project and works to undermine her every chance she gets). A social institution is “more consistent” with a certain form of solidaristic integration when it creates fewer and less powerful reasons for people to distance themselves from one another in the relevant ways. A social institution is “less consistent” when it creates more and more powerful reasons for people to distance themselves from one another in the relevant ways.

According to the fragmentation account, social institutions are sometimes embedded in particular normative relationships. When an institution X is embedded in a normative relationship R, the form of solidaristic integration required by R is an abstract pattern of possible interaction that X must respect, that is, X must be properly consistent with this pattern.²⁴² Proper consistency allows for X to create reasons for distancing, but only so long as (a) these reasons do not bear on the form of integration that matters to the relationship or (b) these reasons are relatively minor. When the reasons for distancing lack relevance and seriousness, the institution is not antithetical to the relevant form of integration and is therefore not morally defective. When the reasons are relevant and serious, the institution is antithetical and therefore morally defective.

7. Three cases

The fragmentation account gives us a way of explaining why engulfing RDAs are sometimes morally defective in virtue of their rivalry defining character. I will demonstrate the force of the view by showing how it can explain the moral defect in three institutional arrangements.

A. Marriage

²⁴² The fragmentation account does not assert (or deny) that the abstract pattern of social relations is a moral standard that citizens can use in directly assessing one another’s conduct as individuals. The important point for my purposes is simply that the pattern is part of a complex moral standard for assessing social institutions.

Imagine that a married couple sets up a household. As part of their life together, they play *Scrabble*. The game starts out as a relatively harmless diversion, but after a while, they start playing more and more often, and they start betting on the outcome. At first, it's small stuff: who will wash the dishes after the game. But slowly, the stakes get higher—more money, more bragging rights—until eventually, winning and losing determine fundamental issues, such as who will give up their career if they have children. When the stakes get extremely high, the institution pits the married partners against each other in a profound way: when they are playing—and they are almost always playing—each partner must effectively undermine projects that are fundamentally important to her partner (e.g. her partner's career) in order to keep winning and prevent the damage from being done to her own projects.

The fragmentation account can explain the moral defect in high-stakes *Scrabble*. Marriage is a normative relationship that requires people to care for each other in certain ways, including by sharing in one another's career aspirations. The rules of the household are properly understood as being embedded in the marriage relationship, so the form of solidaristic integration required by the relationship constrains the rivalry defining character of the household. As a contained and low-stakes affair, *Scrabble* creates reasons for distancing, but the reasons are neither relevant nor that serious, so the institution is not antithetical to marital solidarity (i.e. a "friendly competition"). But when *Scrabble* becomes an all-encompassing, high-stakes affair, it creates powerful reasons for each partner to act with an extreme form of disregard for her partner's aspirations. Here the game crosses the threshold from an acceptable RDA to a morally defective one.

B. Physics department

Next consider the original physics department example. Recall that under S1, junior faculty members have "clear pathways" to tenure: they can each formulate and successfully carry out a plan to get tenure without undermining the other's prospects. But under S2, one junior faculty member's formulating and successfully carrying out a plan to get tenure would effectively block the other from doing the same thing.

The fragmentation account can explain the distinctive moral defect in S2. The departmental relationship requires a form of solidaristic integration among faculty members, i.e. professional collegiality, which involves members sharing in one another's professional projects and aspirations. The department's tenure scheme is properly understood as being embedded in this relationship, so the form of integration required by the relationship constrains the rivalry defining character of the institution. S1 does not pit anyone against each other, so it does not create any reasons for distancing. S2, however, creates powerful reasons for distancing, as each junior faculty member must act with an

intense form of disregard for the other's professional projects and aspirations in order to get tenure herself. According to the fragmentation account, the moral defect in S2 is that the arrangement is embedded in the departmental relationship, but it has a rivalry defining character that is antithetical to professional collegiality.

Note that the fragmentation account would allow for departmental practices that pit faculty members against each other within certain limits. For example, suppose that the department allocates research funding in such a way that each faculty member gets a certain share in a fixed pool of total funding depending on her position in a rank ordering of members based on their recent publications. Although the funding scheme creates obstacles to collegiality, these obstacles are acceptable so long as the differences in people's track records do not make too big a difference to their absolute levels of funding. When the stakes lie within the proper bounds, the funding system is engulfing and rivalry defining, but remains an acceptably "friendly" or "collegial" competition.

C. Liberal democracy

Consider now the case of a labor market with potentially life or death consequences. Suppose that a liberal democracy has a publicly funded health care system, and is considering moving to an arrangement in which it would be up to each citizen to secure a market income and buy private health insurance for herself and her dependents. Under the high stakes arrangement, labor market participants must each follow a course of action that will block some other citizen and her dependents from getting health care in order to secure health care for herself and her dependents. Even in simply holding on to their jobs, citizens must act with an intense disregard for the aspirations of the unemployed who are trying to break into the workforce and get access to health care.

The fragmentation account can explain the distinctive moral defect in a high stakes labor market. The major institutions in a liberal democracy are properly understood as being embedded in the political relationship. This relationship requires solidaristic integration among citizens with respect to one another's pursuit of the common good. Under the first arrangement, society's basic institutions do not create an institutional arena in which one citizen's actions could affect another citizen's institutional prospects for securing health care. Under the second arrangement, society's basic institutions create just such an arena. In pursuing a job under the second arrangement, a citizen *A* may damage some other citizen *B*'s effort to secure health care for *B* and her family, but *A* must disregard the potential impact of her actions on *B* in order for *A* to secure health care for *A* and her family. According to the fragmentation account, the moral defect in the high stakes labor market is that it is an engulfing

arrangement embedded in the political relationship, but it creates reasons for distancing that are both relevant and serious. It is therefore antithetical to civic solidarity.

8. Extending the account to political life

I have presented the central idea in the fragmentation account. This account provides a general explanation for why social institutions are sometimes morally defective in virtue of their rivalry defining character. The fragmentation account applies to all kinds of institutions, but my main interest is with the large-scale social institutions that are the central concern of political philosophy. In order to develop the fragmentation account as a political concern, and to forestall certain objections, I want to show how the account can fit into a liberal conception of political morality.

On my view, the correct principle for assessing the rivalry defining character of engulfing institutions in a liberal democracy is the *Containment Principle* (CP). This principle is meant to work alongside other principles, such as Rawls's first principle of justice or Mill's principle of liberty in an overall account of political morality. Much like these other principles, the CP can serve as part of a political conception of justice, one that can be endorsed from the perspective of many reasonable comprehensive doctrines. The principle reads as follows.

Containment Principle (CP): Aspects of a social order that are not essential to basic liberal freedoms must be arranged so that the sphere of competition in society does not create reasons for distancing that are antithetical to solidaristic integration with respect to the common good.

The most important feature of the CP is that it recognizes the priority of liberal freedom. The CP assumes that social institutions must put individuals in a position where they can each form, revise and pursue their own particular conceptions of the good. Among other things, this means that basic institutions must provide people with what Rawls (1993a) calls a "fully adequate" scheme of civil and political freedoms, including the freedom of thought and expression, freedom of religion, freedom of association, freedom of movement, and so on. The CP also takes it for granted that when people have

liberal freedoms, they will each find their way to distinct and sometimes incompatible conceptions of the good.²⁴³

What the CP stresses, however, is that a social order that provides citizens with liberal freedom will also create an institutional arena in which citizens exercise this freedom. This arena may be competitive. For example, individuals in a liberal democracy may exercise their liberal freedom against the background of a competitive system of civic honors that pits them against each other. Or they may exercise their liberal freedom against the background of an economic system that requires them to compete for basic necessities. The profound importance of liberal freedom will justify those rivalry defining features of the institutional order that are essential to this freedom, but it will not justify rivalry defining elements that are not essential to it. So the CP requires that a liberal democratic social order must moderate those aspects of its rivalry defining character that are not essential to liberal freedom.

We can think of the institutional measures that answer to the CP as falling into two basic categories.

A. Restricting the scope of the competitive sphere.

To satisfy the CP, a liberal democratic social order may favor noncompetitive mechanisms for distributing goods that are part of the common good. For example, institutions can assign goods unconditionally (e.g. civil liberties, basic income), assign goods according to need-based criteria (e.g. Medicaid), or assign goods using lotteries (e.g. organs). Institutions can also use noncompetitive tests—e.g. all those who get at least an 1100 on the SAT will be guaranteed a spot in a public or private university (more on this below). Using noncompetitive mechanisms would shrink the competitive sphere, eliminating institutional possibilities for one citizen's actions to affect another citizen's pursuit of some aspect of the common good.

B. Lowering the stakes in the competitive sphere

²⁴³ If there are many different ways of life that are all worthwhile but not mutually compatible, then any social order that protects liberal freedom will be an RDA in my sense.

In many cases, a liberal democracy will have good reason to use competitive arrangements to distribute goods that are part of the common good. For example, a market for housing has certain advantages over bureaucratic assignment because a market (when properly managed) can do a better job of matching houses to people's needs. The drawback is that each citizen typically finds that in order to get the house that she wants, she must formulate and carry out a plan that will effectively block others from getting the houses that they want.

To satisfy the CP, a liberal democratic social order can lower the stakes when elements of the common good are involved. In the case of the housing market, if there are fewer decent homes available, market participants pose a more serious threat to one another because they may deprive one another of decent housing. But with more decent homes available, market participants would pose a less significant threat because they would be less likely to deprive one another of decent housing. A liberal democratic social order could lower the stakes by increasing economic capacity and directing this capacity to producing the relevant goods.

Surveying the world today, many actual liberal democracies—e.g. Denmark, Sweden, Norway, and Canada—have institutional arrangements that incorporate many of the measures just described. These societies use unconditional assignment, need-based assignment, lotteries, and noncompetitive testing to shrink the sphere of competitive distribution. And these societies use markets to increase output and redistributive measures to diffuse this output so as to lower the stakes in the competitive sphere. On my view, these societies are more “social democratic” as compared to the United States, not simply (or even primarily) because they generate a more *equal* distribution of resources. They are more “social democratic” because they define a more limited and more moderate sphere of competition among citizens.

9. Understanding the moral significance of rivalry

Turn now to an objection. Suppose that a liberal democracy uses the measures discussed in the last section to moderate and contain the sphere of competition in society. The institutional order now satisfies the requirements of the CP. Suppose that it is possible to reform the order so as to increase aggregate output and give everyone a more extensive

array of resources to use in pursuing their conceptions of the good. The only issue is that the reform would violate the requirements of the CP. Some readers might object that it would be irrational for society to forego the benefits of the reform simply because this would involve pitting people against each other more intensely.

The first thing to stress in response to the objection is that political morality is not fundamentally about welfare or well-being. Consider that a preferential hiring scheme that excludes women or minorities from positions of authority in society might actually increase economic output and generate more resources for all citizens to enjoy, even those in the excluded groups. Nonetheless, an arrangement along these lines would be morally defective. The reason is that political morality is not primarily about the welfare that individuals enjoy or their level of well-being; it is about how they relate to one another. By excluding people from positions of authority, the preferential hiring scheme publicly marks certain people as inferior and gives them a reasonable basis for rejecting the arrangement. This means that the scheme is inconsistent with a way of citizens relating to one another in which they can justify their conduct to each other in terms that are properly respectful of each person's distinct point of view (see Scanlon 1998; Rawls 1999; Rousseau 1997a).

The requirements of political morality are often best understood in terms of a way of citizens relating to one another that is rooted in the possibility of mutual justification among persons. But some requirements of political morality are best understood in terms of a way of citizens relating to one another that is rooted in a more concrete relationship. Many philosophers (see p.4 above) recognize that the requirements of political morality are often best understood as the requirements of a special relationship that citizens stand in with their fellow citizens. For example, Ronald Dworkin (1986) argues that the obligation to obey the law in a liberal democracy is best understood as an "associative obligation" that citizens owe one another as partners in an ongoing political relationship. Similarly, Rawls (2005b) argues that the requirements of public reason apply to citizens insofar as they stand in a political relationship, where these requirements "specify the nature of the political relation in a constitutional democratic regime as one of civic friendship" (447).

The fragmentation account begins with the widely held idea that the requirements of political morality are often the requirements of an ongoing political relationship. But it differs from the views of Rawls, Dworkin, and others, in the way that it interprets these requirements. According to the fragmentation account, the political relationship requires a form of solidaristic integration among citizens, not unlike the integration required among neighbors or professional colleagues. Social relations should ideally take the form of citizens giving one another's aspirations for liberty, income, health care, and so on, a status in their reasoning that is similar to the status that they give to their own aspirations for these goods in their reasoning. More importantly, the political relationship prohibits citizens from adopting arrangements that are antithetical to this form of solidaristic integration. A high stakes labor market would interfere with citizens sharing in one another's aspirations in the right way, and because it tears at the fabric of affirmative relations that exist or could exist among us, the political relationship prohibits this arrangement.

A fundamental advantage of the fragmentation account is that it avoids the dualistic view of social relations that is implicit in the theories of Rawls and Dworkin. In Rawls's case, citizens stand in a political relationship that is so intimate that they must each tailor their public arguments and voting behavior to be properly sensitive to the innermost religious and political convictions of their fellow citizens (i.e. public reason). Yet *these very same people* stand in no significant political relationship in economic life, so there is nothing especially problematic about a market arrangement insofar as it may require citizens to foreseeably smash their fellow citizens' livelihoods in order to avoid having their own livelihoods smashed. I find this dualistic view hard to believe: if we are civic friends, then we are civic friends, whether in economic life or in political life. A fundamental attraction of the fragmentation account is that it unifies our view of the institutional morality of economic and political life in one coherent conception of social relations among citizens.

Some philosophers might defend Rawls's view against the charge of dualism by arguing that citizens stand in a political relationship, both in economic life and in political life, but that the relationship is much thinner than friendship. The relationship requires only that citizens relate to one another as fellow participants in a rule-governed enterprise, where participants must be able to justify the rules to one another by appeal to shared principles of the right kind. Understood in this way, the political relationship is more like the internal relationship among players in a game: players must ensure that the rules are fair and can be shown to be fair, but there is no background requirement to share in one another's aspirations. A view along these lines would allow for intensely competitive arrangements in economic life but would avoid the charge of dualism.

The problem with the response is that in avoiding the charge of dualism, it leaves us with a conception of the political relationship that is just *too* thin. After all, if the political relationship requires only that people relate to each other as participants in an enterprise with mutually justifiable rules, then even enemy soldiers on a battlefield may live up to its requirements: soldiers may relate to one another as fellow participants in an enterprise with rules that they could justify to one another, e.g. no chemical or biological weapons, benevolent quarantine for prisoners, etc. But combatants trying to destroy each other's basic projects and capacities are not living up to the requirements of the political relationship, even if they regulate their behavior according to a shared sense of justice in combat.

The fragmentation account offers a more compelling picture. Citizens must relate to one another as fellow participants in a mutually justifiable, rule-governed enterprise. But at a deeper level, they must also relate to one another as partners in a relationship that requires them to share in one another's aspirations for the common good. In effect, the rules of civil life are supposed to be the rules for a system of *good will* among citizens—i.e. civic friendship. Adding solidaristic integration gives us a richer conception of the political relationship, a conception that distinguishes the mutually justifiable rules appropriate to a political community from those that might be appropriate to enemy combatants or to people whose only relationship is that they are players in the same game.

10. Is the view too demanding?

Consider finally an objection related to social costs. Most any principle of political morality may interfere with the realization of other social objectives, but some readers might argue that the costs of conforming to the CP would be so high that the principle could not be part of a plausible conception of political morality.

Here I want to stress that the CP is not an egalitarian principle, so it is compatible with significant performance incentives. In one kind of case, institutions may create incentives when the rules do not situate people in such a way that one person's securing the benefit would interfere with another person's securing it. For example, many liberal democracies today create incentives for high school students to develop their talents by administering standardized tests in different subject areas, where a combined score above a certain threshold earns a student a place in a university. The incentive scheme is structured so that most students will not meet the threshold, but anyone who does meet it will get a spot in a university, regardless of how other students do. The scheme incorporates funding rules that require governments to set aside the equivalent of a certain dollar amount of funding per student using projections based on the size of the college age cohort in the population. Any student who passes the test will carry this dollar amount with her to whatever college she attends. The rules also require governments to set aside more funding from tax revenues if this becomes necessary in any given year. An arrangement along these lines situates students so that one student does not get into university by keeping another one out, and so the performance incentives involved do not interfere with students (or their families) sharing in one another's college aspirations.

On the other hand, the CP does put limits on competitive incentives. Maybe these limits would have dire consequences? Here there is good reason to be skeptical. Our culture endlessly glorifies competition, but for every person who thrives in competitive circumstances, there are others who can never do their best because they appreciate how their actions damage others. Competition can encourage technological innovation, but the most fundamental scientific and technological breakthroughs typically develop in universities and other collaborative environments that insulate people from various competitive pressures (Lester and Piore 2004, among many others). Competition can lead to socially wasteful investment in strategic activities (Frank 2011; Frank and Cook 1995), and competition can fragment society in ways that increase the dangers of authoritarian government (Tocqueville 1969). But the most obvious response to the objection is to point to the experience of many European social democracies, which maintain high levels of achievement in scientific, cultural and economic activities, while also moderating and containing the sphere of competition in social life.

11. Conclusion

Large-scale social institutions are very different from ordinary games: they are wide-ranging, unavoidable and have profound effects on people's lives. The competitive character of these arrangements is much more problematic than the competitive character of the typical game of *Scrabble* or tennis. Political morality puts limits on the rivalry defining character of large-scale institutions in a liberal democracy, limits that cannot be understood simply in terms of good feelings, fairness or equality. The fragmentation account provides a compelling explanation of the relevant constraints, and the CP shows how the underlying concern fits into a broadly liberal conception of political morality.

References

- Appelbaum, A. 1999. *Ethics for Adversaries*. Princeton: Princeton University Press.
- Cohen, G. A. 2009. *Why Not Socialism?* Princeton: Princeton University Press.
- De Tocqueville, A. 1969. *Democracy in America*, ed. J.P. Meyer. New York: Harper & Rowe.
- Dworkin, R. 1986. *Law's Empire*. Cambridge: Harvard University Press.
- Frank, R. 2011. *The Darwin Economy*. Princeton: Princeton University Press.
- and Cook, P. 1995. *The Winner-Take-All Society* (New York: Penguin).
- Gomberg, P. 1995. "Against Competitive Equality of Opportunity" *Journal of Social Philosophy* 26/3: 59-74.
- Kant, I. 1996. "The metaphysics of morals" *Practical Philosophy*, ed. A. Wood. Cambridge: Cambridge University Press.
- Lester, R. and Piore, M. 2004. *Innovation*. Cambridge: Harvard University Press
- Marx, K. 1964. *The Economic and Philosophical Manuscripts of 1844*, trans. M. Milligan. New York: International Publishers.

- McMahon, C. 1981. "Morality and the Invisible Hand" *Philosophy and Public Affairs* 10/3: 247-77.
- Rawls, J. 1999. *A Theory of Justice*, revised ed. Cambridge: Harvard University Press.
- 2005a. "The Basic Liberties and their Priority" in *Political Liberalism, expanded edition*. New York: Columbia University Press.
- 2005b. "The Idea of Public Reason Revisited" in *Political Liberalism, expanded edition*.
- Rousseau, J. 1960. "Letter to M. D'Alembert on the Theatre" in *Politics and the Arts*, trans. A. Bloom. Ithaca: Cornell University Press.
- 1979. *Emile*, trans. A. Bloom. New York: Basic Books.
- 1997a. "On the Social Contract" in *The Social Contract and Later Political Writings*, ed. V. Gourevitch. Cambridge: Cambridge University Press.
- 1997b. "Considerations on the Government of Poland" in *The Social Contract and Later Political Writings*.
- Scanlon, T. 1998. *What We Owe to Each Other*. Cambridge: Harvard University Press.
- 2003a. "Due Process" in *The Difficulty of Tolerance*. Cambridge: Cambridge University Press.
- 2003b. "The Difficulty of Tolerance" in *The Difficulty of Tolerance*.
- 2008. *Moral Dimensions*. Cambridge: Harvard University Press.
- Walzer, M. 1983. *Spheres of Justice*. New York: Basic Books.

CHICAGO ATTRACTIONS

John Hancock Tower:

The best-kept secret in Chicago tourism is the Signature Lounge, located on the 96th floor of the Hancock Tower, 875 N. Michigan Ave. This bar/restaurant provides guests with a 360 degree view of Chicago and Lake Michigan for the price of a drink --there is no admission fee.

The Magnificent Mile:

Chosen as one of the ten great avenues of the world, the Mag Mile is located just north of the loop and is Chicago's most prestigious shopping district. Water Tower Place, a very large mall, is located at 835 N. Michigan Avenue. Walking south on Michigan Ave (or taking any of the many buses) you will end at the Wrigley Building down on the river (which you can follow into the loop and to Millennium Park and the Art Institute).

Chicago Architecture Foundation Boat Tour:

\$44 for daytime cruises and \$46 for nighttime cruises, 90 minutes long. Dock location is southeast corner of the Michigan Avenue Bridge and Wacker Drive. Look for the blue awning marking the stairway entrance. You can buy tickets online.

Millennium Park:

Millennium Park is located in the heart of downtown Chicago. It is bordered by Michigan Avenue to the west, Columbus Drive to the east, Randolph Street to the north and Monroe Street to the south. This park is open daily from 8am to 11pm. Admission is free. Attractions include the enormous mirror-surfaced bean sculpture, the Cloud Gate bridge, the Crown Fountains, the outdoor amphitheater, and the Lurie Garden.

Shedd Aquarium:

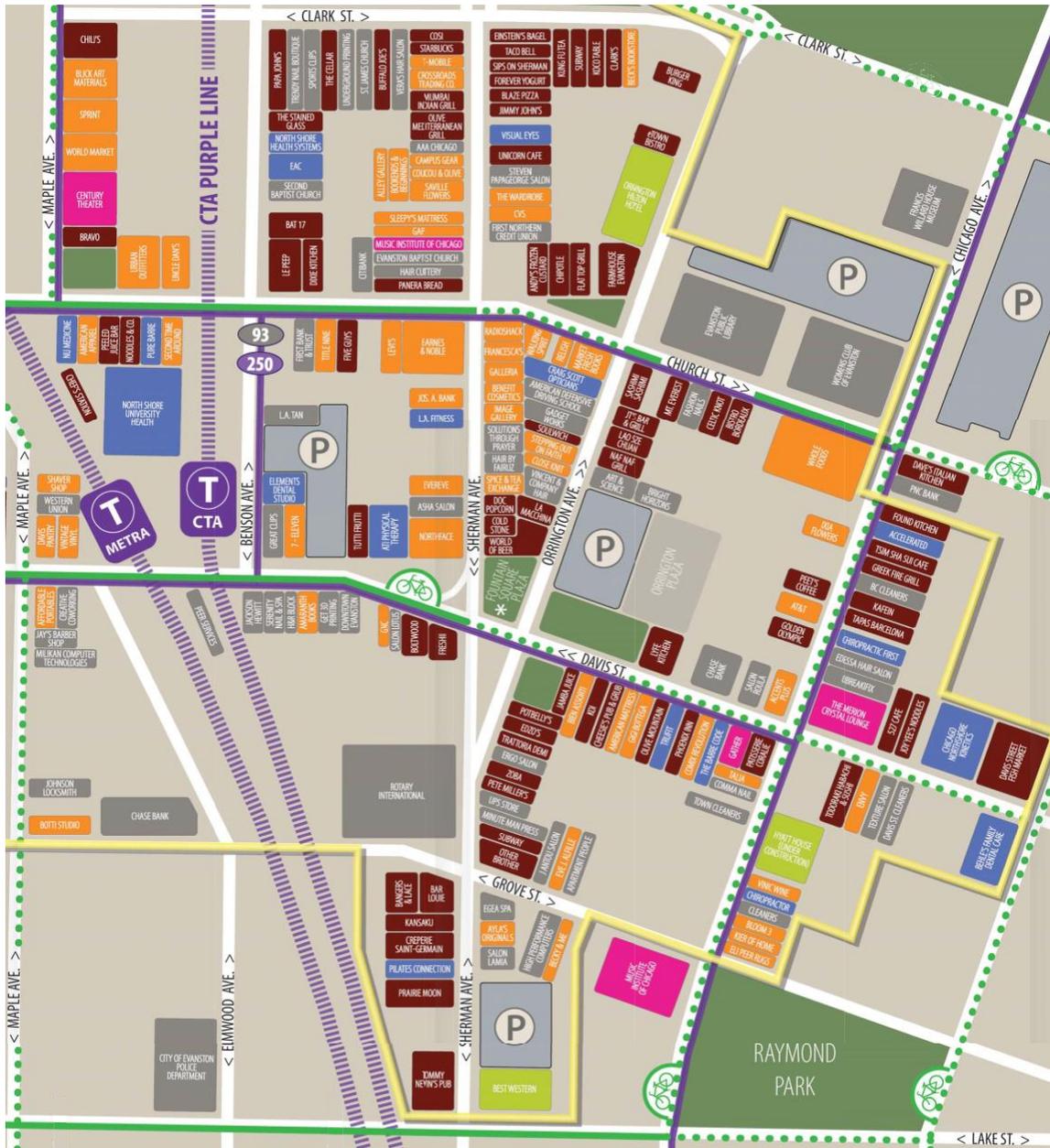
Museum Hours: Weekdays: 9am-5pm & Weekends: 9am-6pm. Admission: \$8 adults for aquarium only, \$31 for all-access pass that includes Oceanarium, Wild Reef, Amazon Rising, the Caribbean Reef, Waters of the World, and others. To get to the museum, take the red line L to the Roosevelt stop and board a museum trolley or take the #12 bus.

The Field Museum:

Museum Hours: 9am-5pm. \$38 for an all-access pass. Take the red line L to the Roosevelt stop and board a museum trolley or take the #12 bus.



Downtown Evanston



EAT/DRINK

SHOP

BE ENTERTAINED

STAY